DOCUMENT RESUME

ED 238 651 RC 014 528

AUTHOR	Bhola, H. S.
TITLE	Evaluating Development Training Programs. (A Practical Manual for Trainers of Development Workers and Change Agents in the Third World).
INSTITUTION	German Foundation for International Development, Bonn (West Germany).
PUB DATE	82
NOTE	293p.
PUB TYPE	Guides - General (050) -- Books (010)
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	*Change Agents; Community Development; Data Analysis; Data Collection; Definitions; Developing Nations; Evaluation Criteria; *Evaluation Methods; Foreign Countries; *Improvement Programs; Material Development; Models; Professional Training; *Program Evaluation; Rural Development; *Staff Development; Technical Writing; *Training Methods; Training Objectives
IDENTIFIERS	*Development Education

ABSTRACT

A manual for trainers of development workers and change agents in the third world provides guidelines for evaluation of development training programs. Chapters define evaluation, present a general model of development, and discuss evaluation in development training programs. Designing training systems for development workers and change agents, asking evaluation questions, making evaluation agendas, writing formal proposals for evaluation studies, and monitoring and quick appraisals are covered in subsequent chapters. Naturalistic and scientific paradigms of evaluation, concept analysis, change indicators, quality standards, and evaluation tools and instruments are discussed. Other chapters detail product evaluation (pre-testing of instructional and training materials), designing evaluation studies, and the concept of design in naturalistic evaluation. Implementing evaluation studies is discussed in connection with problems of data collection in the real world. Processes and techniques of data analysis and statistics for evaluators of development training programs are covered. Guidelines for writing evaluation reports are given. A final chapter discusses politics of evaluation and evaluation standards. An appendix covers history and use of the Action Training Model. A bibliography presents sections on training design (26 entries), evaluation models and techniques (65), and evaluation of training and development (31). A glossary defines relevant terms. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
Josef Muller

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

EVALUATING DEVELOPMENT TRAINING PROGRAMS

(A Practical Manual for Trainers of Development Workers and Change Agents in the Third World)

By

H. S. Bhola

Indiana University

1982

DEUTSCHE STIFTUNG FUR INTERNATIONALE ENTWICKLUNG

German Foundation for International Development

To my wife, Joginder --

One tough evaluator!

TABLE OF CONTENTS

Page

Preface 5

CHAPTER I: What is evaluation? 15

II: A general model of development 22

III: The promise of evaluation in development training programs 27

IV: Designing training systems for development workers and change agents 32

V: Asking evaluation questions, making evaluation agendas, and typical evaluation themes 54

VI: Writing formal proposals for evaluation studies 71

VII: Monitoring and quick appraisals 87

VIII: Models of evaluation 93

IX: Concept analysis, indicators of change and standards of quality 115

X: Evaluation tools and instruments 145

XI: Product evaluation or pre-testing of instructional and training materials 168

XII: Designing evaluation studies, and the concept of design in naturalistic evaluation 177

XIII: Implementing evaluation studies: problems of data collection in the real world 193

XIV: The process and techniques of data analysis 198

XV: Basic statistics for evaluators of development training programs 226

XVI: Writing evaluation reports 242

XVII: Politics of evaluation and evaluation standards 253

APPENDIX I: The Action Training Model 262

Bibliography 272

Glossary of terms 282

PREFACE

This is a book on the evaluation of development training programs -- courses, camps, seminars and workshops organized to train development workers.

Development is on the national agendas of almost all of the Third World countries today. To teach farmers, workers and housewives in villages and cities, new knowledge, new attitudes and new skills, hundreds of thousands of development workers (or change agents) are being trained all over the Third World.

This training is being conducted through ad hoc training programs; in training centers and institutes specially set up for the purpose; or in universities and other institutions of higher education. Training objectives vary from the economic to the social and the political. Training philosophies and methodologies differ widely. The levels of personnel trained range from front-line workers to specialists, planners and managers at district, regional and central levels. Both civil servants and volunteers may be covered. Training experiences may vary in duration from half a day to a couple of years.

Accumulated experiences in the training of change agents varies, of course, from country to country. Some developing countries have already built extensive networks of institutions to meet the training needs of development workers. India in Asia, Kenya in Africa and Brazil in South America could be cited as examples of such countries. Other countries have not been as fortunate. They have been constrained by lack of resources and still have a long way to go in fulfilling their development training needs.

A mood of stocktaking of training experiences, however, seems today to be widespread. Countries that already have gained long experiences in the training of change agents want to evaluate their experiences. Others who are just beginning, want to make sure that they do not repeat the mistakes that the pioneers may have made before them. Within the development training sector, those currently engaged in development training as trainers or as administrators wish to be able to evaluate training objectives and training methods and materials, to maximize impact on their trainees, and ultimately on the communities undergoing development. Evaluation of the effectiveness of development training is, therefore, today an area that is both quite popular and inherently significant.

This book is addressed to the trainers of development workers and change agents for various development sectors, such as adult basic education and adult literacy, agricultural extension, family planning, health and nutrition education, cooperative education, community development, social welfare, development administration, leadership training, development support communication, integrated rural development, trade union education and political education. Focus is on the Third World -- the areas of poverty in Africa, Asia and Latin America. This particular choice of audience, their educational background and experience, and the conditions of their work have determined the content, the general approach and the style of this book. While development trainers are the primary audience of this book, those engaged in the planning and administration of development programs in national and international organizations might also find the book of some interest.

The five introductory chapters provide a framework for the discussion of methods and techniques of evaluation included in later chapters. The first chapter of the book states the essential purpose of conducting evaluation. In a second chapter, a general model of development is presented. The model points up the significant role of human resource development within the overall developmental processes and puts the need for training development workers and change agents in the proper perspective. The returns that will most likely accrue from the evaluation of development training programs are then weighed and considered.

In the fourth chapter of the book, some preliminary ideas on training design are offered. A model for developing useful descriptions of various training programs is then included, on the premise that to evaluate a program, the evaluator must first be able to describe it in conceptual terms. This chapter should enable readers to look at training programs as training systems, with particular structures and processes within, and particular interfaces and interconnections without. The model presented in this chapter may also provide norms and standards for making evaluative judgments in regard to training programs being evaluated. Finally, in the fifth introductory chapter evaluation questions of significant interest to development trainers are generated; and the criteria for the choice of training issues for evaluation are discussed.

In the chapters that follow (Chapters VI to XVIII), the whole array of evaluation topics has been discussed including writing evaluation proposals, monitoring and quick appraisals, evaluation models, concept analysis, development of indicators, evaluation design, construction of tools and instruments, product evaluation, data collection and data analysis, basic statistics and dissemination of evaluation results.

All this content has been included in this one book because all these topics do go together. Also, because this may be the only book our readers may have access to or may read for some time. The book presents what we hope is a judicious mix of the theoretical and the practical content. Topics such as the development model, training design and evaluation models, by necessity, were discussed at a highly conceptual level. All other topics have been discussed at a much more practical level.

Evaluating Development Training Programs is meant to be a first systematic introduction to the evaluation of development training programs. This is not a book for the practicing evaluator, nor do we seek to prepare evaluation experts. Our aspirations are more modest. We seek to prepare "trainers-evaluators" -- professional trainers who have acquired a sufficient level of evaluation skills to be able to develop needed feedback for the improvement of their development training programs.

The emphasis is on internal evaluation, evaluation handled by the trainer and often relating to the trainer's own program. We promote low-cost evaluation and suggest that evaluation of development training be handled as far as possible from within the resources already available to the training program or to the training institution.

Both formative evaluation (evaluation conducted during the formation of the program with the intent to improve its design), and summative evaluation (evaluation conducted at the end of a program to sum up its results) have been covered. Both the scientific (quantitative) and the naturalistic (qualitative) paradigms have been discussed. Our methodological orientation is pragmatic, even opportunistic. We believe that some feedback on training effectiveness is better than none at all; and if the best of evaluation design and measurement is not possible in a situation, the second best is all right too.

The language used in the book has been as simple as possible. Where technical terms had to be used, they have been explained in the text, and more elaborate definitions have been included in the Glossary at the end of the book.

The book does not promise miracles, however. It will have to be read and read carefully by trainers-evaluators. Some chapters and portions of other chapters may need to be read more than once for maximum benefit. References to the dictionary or to the Glossary of terms may be necessary. Suggested applications of some of the concepts will need to be actually worked out; and statistical tests must be learned by applying them to available data. In some cases, additional reading will be called for.

When used as a text for seminars and workshops of short duration, the various chapters of the book need not be read in the order in which they have been presented in the book. Both the content and the organization of content for the workshop should be developed participatively, with the particular workshop group, in terms of the particular needs and interests of that group, at that particular time. Thus, one workshop may start with how to develop an evaluation proposal and go through writing evaluation questions, to developing indicators, to constructing questionnaires and structured interview schedules, to the analysis of data.

Another workshop may start with descriptions of training systems and go through developing indicators, to developing questionnaires to basic statistics. Yet another workshop could start with the models of evaluation and, in the choice of other content, may emphasize naturalistic evaluation strategies.

In each case, workshop organizers should point out clearly to participants what will be read from the text, in what sequence, by what time, indicating page numbers for each assignment. Participants should, at the same time, be encouraged to read the whole book from the beginning to the end. This should be done during the workshop if at all possible; or immediately after the workshop. It is important that trainers-evaluators get a sense of the total picture of evaluation and do not confuse a part with the whole. When carefully read and understood, the book will teach not only evaluation of development training but also a lot about training design and about development planning itself.

This book has emerged from a series of workshops and panels on evaluation of basic education and development training programs organized by the German Foundation for International Development (DSE) in Kenya during 1979-82.

This workshop series had its antecedents in a 2-week workshop on the evaluation of "functional literacy programs" organized in Mombasa, Kenya during May 9-21, 1977 under the joint sponsorship of DSE and the International Institute for Adult Literacy Methods (IIALM). The focus of this workshop, understandably, was on literacy evaluation; and it was attended only by the literacy and adult education officers of the then division of adult education of the Ministry of Housing and Social Services of the Government of Kenya.

It was at the 1977 Mombasa workshop that the idea of organizing evaluation workshops for "adult educators" from all the various development sectors was born. The realization emerged that there were adult educators in other development sectors such as agriculture extension, health and nutrition, family planning, cooperatives and workers education; and that a diffusion of evaluation skills within this expanded adult education culture will have a salutary effect on overall development plans and development outcomes. At the same time, there was a sharpening of focus on teaching evaluation skills to trainers in training centers and institutes engaged in preparing development workers and change agents. It was hoped that this strategy would maximize the effects of the evaluation training offered in Kenya. Trainers receiving training in evaluation techniques will improve delivery of their own training. They will also be able to share some of their evaluation skills with development workers they will, in turn, train themselves.

Thus, emerged the series of workshops and panels on the evaluation of development training programs, and with it the Action Training Model. The Action Training Model, which has been described more fully in Chapter IV, included the features of participative planning; development of evaluation proposals during the workshops and their implementation back at the training center; provision of technical assistance in short panels and continuously through the establishment of an Educational Evaluation Resources Committee; and emphasis on developing local capacity to later continue the project without technical assistance from outside.

With the evaluation workshop of June 1982, the first phase[1] of this series of workshops and panels comes to an end, and a second phase begins. In a second phase of four years, 1982-86, the series of workshops and panels on the evaluation of basic education and development training will be continued by the Institute of Adult Studies (IAS), University of Nairobi, Kikuyu, Kenya. The Action Training Model will be further implemented. The local organizers will receive some funding from DSE but no outside consultants will be used.

This book, as we indicated earlier, was born in the context of the workshop series just described. In the first workshop during May 21 to June 1, 1979 at Kericho, Kenya, H. S. Bhola's Evaluating Functional Literacy (Hulton Educational Publications, Amersham, Bucks, U.K., 1979) was used as the text for the workshop. It became quite clear, however, in using this book that while transfer from "functional literacy" to "development training" was possible, it was problematic in that setting. In a short two-week workshop, it was best not to impose on the trainees the additional burden of making the conceptual transfer from literacy to development training and unduly complicate instructional tasks. The best thing to do was to develop materials that dealt directly with the world of work of development trainers, related with their institutional settings, and their special concerns in training and evaluation. That indeed was what we undertook to do.

---

[1] During the first phase of the DSE program, the following workshops and panels on the evaluation of basic education and development training programs were conducted: (1) Workshop during May 21 to June 1, 1979; (2) Panel during January 4-10, 1980; (3) Workshop during August 18-29, 1980; (4) Panel during March 23-28, 1981; (5) Workshop during June 1-13, 1981; (6) Panel during February 22-27, 1982; and (7) Workshop during June 28 to July 10, 1982. The first five events took place in Kericho, Kenya and the last two in Naivasha, Kenya.

We started with handouts[1] dealing with the immediate and expressed needs of workshop participants, thereby complementing the literacy evaluation text. We are pleased that we are now able to present this complete and comprehensive treatment of the topic of evaluating development training programs for use by development trainers. This new book will find its first use at the June 1982 workshop in Naivasha, Kenya and will continue to be tested in the workshops and panels that follow. We do hope to use the feedback resulting from the utilization of the book for preparing it for later publication.

I have had the pleasure and privilege of directing all of the workshops and panels (with the exception of the Naivasha Panel of February 22-27, 1982) during the first phase of the workshop series. But these things are never really done singly, by oneself. In a very real sense, these workshops and panels have been co-directed by Dr. Josef Muller of the DSE who has been my full partner both in developing the concept of the Action Training Model and in sharing instructional responsibilities at the workshops and panels. Dr. John W. Ryan, now of the Literacy Division, Unesco, Paris (and previously, director, IIALM); Dr. Tom Mulusa of IAS; and Mr. Joseph Dondo, previously of IAS, and now of the Kenya Institute of Management, are the other intellectual parents of the Action Training Model.

---

[1] H. S. Bhola, Designing and Evaluating Development Training Programs: A Working Document (August 1980); Designing and Evaluating Development Training Programs: Part II (March 1981); and Data Displays for Data Analysis: Naivasha Panel (February 1982).

My sincere thanks to all of my colleagues from Kenya who have constituted the faculty for these workshops and served on the Educational Evaluation Resources Committee, especially, Joseph Dondo, Tom Mulusa and Daudi Nturibi of IAS; and Alice Waka of the Kenya Institute of Education. My thanks also to Peter Kinyanjui, director, IAS who has provided insightful leadership and unfailing support to the project; and has now decided to go ahead with the second phase of the workshop series.

Last, but not the least, my thanks and acknowledgements to the trainees of the workshops and panels who were such good learners and such good teachers.

Indiana University
Bloomington, Indiana, U.S.A.

H. S. Bhola

CHAPTER I

WHAT IS EVALUATION?

Evaluation is the process of assigning values to judge the amount, degree, condition, worth, quality or effectiveness of something. As human beings, we are perpetual evaluators. We evaluate ideas, attitudes, actions, characteristics and possessions of ourselves and of others, at home and at work, day in and day out, sometimes self-consciously but often unthinkingly; and make judgments on the basis of those evaluations. If life can be defined as a series of decisions, then being is evaluating.

In recent years, evaluation has emerged as a specialization that teaches us how to make evaluations that are self-conscious, systematic, objective and dependable; and which can, therefore, be more useful in making appropriate decisions -- in our case, decisions about implementing development training programs.

Let us examine a few definitions of evaluation that have appeared in literature over the years:

Professor Lee J. Cronback defined evaluation simply as "the collection and use of information to make decisions about an educational program."

Dr. Marvin C. Alkin described evaluation as the "process of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analyzing information in order to report summary data useful to decision-makers in selecting among alternatives."

Dr. Daniel L. Stufflebeam defined evaluation as "the process of delineating, obtaining, and providing useful information for judging decision alternatives."

In a recent book (See Bibliography), Professors Egon G. Guba and Yvonna S. Lincoln have defined evaluation "as the process of describing an evaluand [the entity being evaluated] and judging its merit and worth."

From the various definitions included above, it becomes clear that evaluation involves collection and analysis of data objectively, systematically and with exactitude, to serve a variety of decision-making needs. These decision-making needs may relate to (1) modification or improvement of a development training program; (2) certification of a training program in its present form; or (3) making decisions about alternatives within the program, or a total alternative to it. In sum, the objectives of evaluation are always utilitarian in nature.

Evaluation and Monitoring

The word monitoring has recently come into use in the literature of program evaluation. To monitor is to check upon an on-going program for flaws or breakdowns, to enable decision-makers to regulate activities and to undertake corrective action.

As in the case of evaluation, monitoring also serves the needs of decision-makers. Monitoring, however, is only a part of evaluation, and not the whole of it. Monitoring seeks to look at the performance data generated by programs in the process of their implementation and cautions decision-makers about the gap between reality and expectation.

Programs can be made more monitorable by introducing proformas, schedules, and progress reports at important points and levels of the program so that the program can generate significant performance data in the very process of its implementation. Monitoring, typically, requires the establishment of management information systems within programs so that information generated by the program can flow to a central point at fixed intervals of time. There it must be aggregated and analyzed, and stored for retrieval and use at other times.

Monitoring and Supervision

Supervision involves some monitoring but, at its best, supervision is an educational process. Good supervision enables the various functionaries in a program to analyze and evaluate their own performance in relation to the program needs, and to learn and grow on the job.

Evaluation and Research

Evaluation and research are two different professional activities, though the two get often confused. Confusion occurs because the evaluator and the researcher share inquiry designs, methodologies, tools and instruments, and have similar concerns with the validity and the reliability of their findings. Quite often the same one person may be acting as evaluator and researcher at the same time. Evaluation and research, however, differ significantly in terms of their frameworks and their objectives, as the following table should indicate:

Table 1. The different orientations and objectives of the evaluator and the researcher

Evaluator	Researcher
Policy and planning orientation; seeks to clarify planning alternatives and to improve program performance.	Disciplinary and academic orientation; seeks to advance the frontiers of knowledge in the researcher's chosen discipline.

Loyalty is to a particular program or project; choice of evaluation topics is determined by the information needs of decision-makers.	Loyalty is to a particular academic discipline; choice of research topics is determined by the theory and research needs of the discipline.
The methodological choices are often quasi-experimental or non-experimental; the norm for judging the findings is applicability to the program situation and adaptability to other similar program settings.	The methodological choices are typically experimental or controlled; the norm for judging the findings is generalizability.
Time-frame for the production of results is set by the program.	Time-frame for the production of results is set by the researcher and by the internal logic of the research question.
Professional rewards consist in the utilization of findings by decision-makers and demonstrated improvement in the program implementation.	Professional rewards consist in publication of findings in professional journals and favorable comments by professional colleagues.

The distinction between evaluation and research is not, however, water-tight. There is a lot of bad research around. And there are some really good evaluation studies that produce insightful and generalizable findings and get published in prestigious professional journals. Such evaluation studies may be seen to fit the title of "evaluation research," the phrase used by Carol Weiss.

Internal and External Evaluation

Internal evaluation is that conducted within the program system by program specialists themselves. External evaluation is that conducted by evaluators sent from outside.

It is often asserted that external evaluation is more objective than internal evaluation which is rejected as both subjective and political. On the other hand, program specialists often dread external evaluations which they complain are often hurried, superficial, uninformed and political. There is merit in both positions

External evaluations are by no means inherently objective. On the other hand, internal evaluations can be most objective. Both internal and external evaluations can be highly political. The right approach is to assign appropriate tasks to internal and external evaluations, and to make them to perform complementary roles.

In this book, our emphasis is on internal evaluation which we look at as evaluation for growth rather than for control. By conducting internal evaluation, indeed an evaluation of one's own training performance by oneself, the trainer can perform the essential most task of all evaluation which is to improve performance; and, in the process, can grow as a professional trainer.

Participative Evaluation

Participative evaluation goes beyond the formalism of both internal and external evaluations and conducts evaluation in participation with the publics concerned. Evaluation becomes both educational and liberating. Essentially, participative (or participatory) evaluation is one that is conducted in mutual collaboration by all those engaged in the conduct of a program. At its best, the organizers play a facilitative role while the people being served by particular programs take over. These people determine, through dialogic action, what the evaluation needs are, what information should be collected and how, and what norms and standards should be used to judge success or failure. Participative evaluation is a methodological approach with strong ideological commitment to radical humanism. We will have more to say about it in a later chapter.

Formative and Summative Evaluations

The concepts of formative evaluation and summative evaluation, introduced by Professor Michael Scriven, are important for evaluators to understand.

Formative evaluation examines different aspects of a program as it is being formed. It is used to correct and improve a program during its planning and implementation.

Summative evaluation typically comes at the end of a program or at the completion of a program phase. It is used to sum up the impact of a program on its client groups or communities.

Product Evaluation

Product evaluation may be seen as a special case of formative evaluation when what is being evaluated is an instructional product, such as, a textbook, a set of posters or charts. a film or a simulation-game.

Summary

Evaluation is the process of assigning values to judge the amount, degree, condition, worth, quality or effectiveness of something. As a specialization, evaluation involves the collection and analysis of data objectively, systematically and with exactitude, to serve a variety of decision-making needs. Evaluation can be usefully distinguished from monitoring, supervision and research. Evaluation may be internal or external; or may be participatory in nature. Evaluation may be formative (conducted during the formation of a program) or summative (conducted at the end of a program to sum up results).

Things to Do or Think About

1. Are you clear about the distinction between evaluation and research? Try to explain the difference between the two to a colleague to his or her satisfaction.

2. From your area of work, think of a topic that seems to fit the category of formative evaluation. Think of another topic that fits the category of summative evaluation?

3. What kinds of information must be generated and collected for developing a monitoring system for your training center or program?

CHAPTER II

A GENERAL MODEL OF DEVELOPMENT

If we are interested in the evaluation of development training, we have got to understand all of the three processes involved -- evaluation, training, and development. Therefore, in presenting a general model of the process of development, we are by no means straying away from the main concern of this book.

There are two good reasons why a person preparing to be an evaluator of development training programs should also understand development. One is a general and somewhat abstract reason, but the other is downright practical and functional. Trainers-evaluators engaged in the preparation of development workers and change agents must understand the processes of development to get a sense of the framework within which they work. They must observe the big picture and must understand the interconnections between and among the various development actors and development processes. It is only with such an understanding that their concrete training actions, their training methods and the training effects they obtain, will make sense in the larger perspective of change at the societal level. And, it is such an understanding that will help them to be more than mere technicians or purveyors of skills and information but will make them enlightened and committed change agents -- indeed culture-makers.

A second reason why trainers-evaluators should understand development is more practical and functional. Since training seeks to prepare workers to do better development, it is development that will provide most of the norms and criteria for judging the success or failure of training designs and training methods.

A General Model of Development

A general model of the development process is presented below:

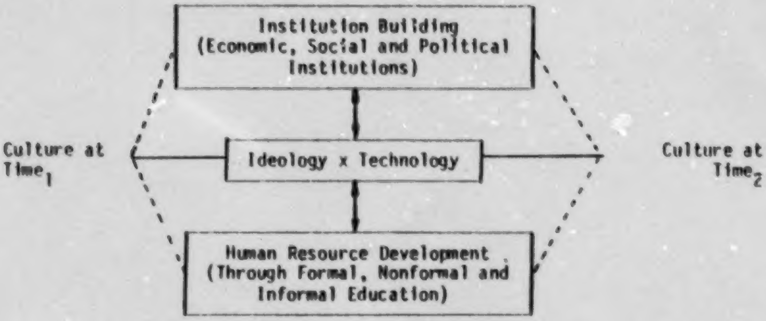

Institution Building (Economic, Social and Political Institutions)
Culture at Time$_1$
Ideology x Technology
Culture at Time$_2$
Human Resource Development (Through Formal, Nonformal and Informal Education)

Figure 1. A general model of the process of development

The process of development and planned change (as a culture goes through some sort of transformation during a particular time period of history), is essentially a calculus of ideology and technology, an adaptation of means to ends. This essential process of interactions between ideology (ends) and technology (means) can be seen to factor out along two dimensions: (1) institution building and (2) human resource development.

While development ideologies may differ somewhat from country to country, a basic core of development is coming to be universally accepted. Development has to be man-centered. Development should involve not merely greater production, but also a just distribution of what is produced. Social and political institutions should be responsive to the needs of the people and should provide opportunities to the people to participate within those institutions.

Technology is what enables societies to implement their ideological visions and their societal aspirations. Technology itself is of two major categories: physical technology which transforms raw materials into economic goods; and social technology which trains and motivates people and invents effective social arrangements.

Two Dimensions of the Development Process

As we have suggested above the "Ideology x Technology" calculus can be seen to occur along two dimensions: institution building and human resource development. Institution building is the name given to the process of improving existing institutions and designing new institutions to undertake the societal tasks that need to be done. Whatever needs to be done systematically and with some continuity needs a system -- an institution of some sort. No wonder, institutions have been called the building blocks of cultures.

The new institutions required by development societies are many. They are economic, social and political: parliaments, primary courts, cooperatives, radio forums, classrooms, rural credit societies, manufacturing plants and factories.

But new institutions will be no more than empty shells if they do not have trained people to man those institutions. That brings into focus the role of human resource development. Human resource development may be undertaken through three different but complementary educational strategies: formal education, nonformal education and informal education.

Formal, Nonformal and Informal Education

Formal education "refers to the hierarchically-structured and chronologically-graded modern 'educational system' that stretches from primary school through the university." Informal education is equivalent to socialization: neither the teacher nor the learner are self-conscious about the process of teaching-learner taking place. Finally, nonformal education refers "to the motley assortment of organized and semi-organized educational activities operating outside the regular structure and routines of the formal system, aimed at serving a great variety of learning needs of different subgroups in the population, both young and old."[1] It is, thus, in nonformal education that the trainers of development workers should be most interested. Trainers of development workers are indeed nonformal educators.

Other Than Training Solutions

The model presented in this short chapter should have helped us put training in a proper development perspective. Training has been shown to be a significant component of development. At the same time, the general model of development presented here should have helped us understand that training is not the solution to all development problems. Problems may lie in defective national visions, in lack of resources of raw materials, in the inadequacy of physical infrastructures, or in the very design of social, economic and political institutions. In these cases solutions will lie elsewhere than in training.

---

[1] The definitions of formal education and nonformal education given here are taken from Manzoor Ahmed and Philip H. Coombs (Eds.). Education for Rural Development: Case Studies for Planners. New York, N.Y.: Praeger Publishers, 1975.

Summary

Trainers-evaluators should be interested in development to understand the true nature of their role as development workers as well as to develop the criteria to judge the success or failure of their training designs. Development basically is a calculus of ideology x technology or the adaptation of means to ends. Another way of defining the development process will be in terms of (1) institution building and (2) human resource development. Human resource development, especially through nonformal education, is of utmost importance in the overall development process. Trainers-evaluators, however, need to be aware that all development problems are not solved by training, or by more and better training.

Things to Do or Think About

1. What are the national aspirations for development in your country? Separate them as ideological (related to ends) and technological (related to means).
2. What kinds of development training needs are being fulfilled in your country by the formal education system?

CHAPTER III

THE PROMISE OF EVALUATION IN DEVELOPMENT TRAINING PROGRAMS

The promise of evaluation in development training programs is rooted in the scope and significance of training in the overall development processes.

The Size and Scope of Development Training

We need to start with the training (or educational) needs of our people. The educational needs of people in the developing world are many and varied. They are political, social and economic; and they are informational, attitudinal and skills-related. Farmers, workers and housewives in the developing nations must learn new developmental values, attitudes and aspirations. Farmers must learn new scientific agriculture and animal husbandry; and new modes of farm management, cooperation and marketing. Workers must learn new technological skills, more productive work habits and patterns, and safety rules at places of work. Housewives must learn new ideas about family care and nutrition. And they should all -- farmers, workers and housewives -- learn to participate in the processes of decision making that govern their lives and work.

But how to make all this possible? Clearly, people can not be left to themselves to learn all this on their own. At the same time, we can not bring them all to school. In fact, we do not have those many schools to accommodate them. Since we can not bring all those people to school, we must take the school to them. And that indeed is what is being done in all kinds of places, in all kinds of settings, all over the Third World through what we have called nonformal education.

To provide this needed education to these new clients and consumers of education, we need new educators. We have called them development workers or change agents or front-line workers. There are other names for them as well: community development workers, agricultural extension workers, village health assistants, barefoot doctors, cooperative education workers, family planning workers, nutritionists, political educators, primary court judges, workers' educators, literacy workers, radio and film forum leaders, discussion leaders and animators. All over the Third World, these new roles are being created to educate the people, to develop the human resources of nations for the ultimate transformation of societies.

The conceptualization and design of these brand new development roles has been one of the most challenging tasks of development work. What should those roles be: agricultural extension worker, cooperatives assistant, family health educator, family planning specialist, political educator, or village level multipurpose worker? How should change agents be recruited to these new roles? What should be their orientation to social change? Should some promote diffusion of knowledge and skills, and some others promote critical consciousness? How should tasks be divided among and between them? If tasks are divided on the basis of specialization, how should integration be achieved in the delivery of services?

The training of these new educators is the most important question of all. How to train change agents for effective performance of their roles? How best to give these change agents the skills they need? How to socialize these change agents in the new norms and values that are necessary for them to internalize to serve the disadvantaged?

The above by no means exhausts all the training needs of development. The trainers of front-line workers must themselves be trained to perform their training tasks, and so on up the line. The following simple diagram should help us visualize the extent and scope of the training problem:

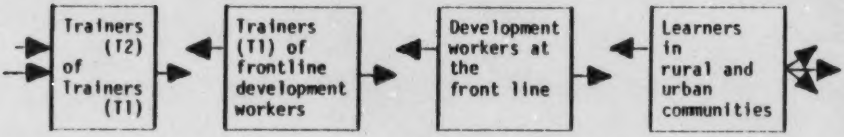

Trainers (T2) of Trainers (T1)
Trainers (T1) of frontline development workers
Development workers at the front line
Learners in rural and urban communities

Figure 1. The trainer-trainee chain in the human resource development approach to development.

The formal education system has been unable to handle the training tasks necessary to implement the total human resource development approach. Only Type T2 trainers have typically come from formal institutions of higher learning -- sociologists, anthropologists, educators, media specialists, etc. -- who, at their best, had acquired experience in development by working in the field on various development projects; and, at their worst, never had any touch with the development world and were discipline-oriented in their training designs.

Type T1 trainers have been trained, as could be expected, outside the formal education system in specialized institutions. Such institutions have proliferated all over the Third World and are now consuming quite an impressive proportion of national education and developmental resources. The Directory of Adult Education Institutions in Kenya (A Guide on Schools, Centers, Colleges, Courses and all Institutions Catering for Continuing Education in Kenya), issued by the Board of Adult Education, Nairobi, 1976, lists a variety of training institutions that include: institutions of public administration, business management, and development administration; of agricultural extension training, animal health and husbandry; institutes and centers for daily management, meat packing, tea, and forestry; cooperative colleges and centers for water development; institutions for the training of workers in health and family planning; for home economics and nutrition; resettlement of cripples; institutions for training personnel for literacy centers, community centers, village polytechnics, vocational centers, farmers training centers; and centers for training personnel for evening schools, correspondence colleges, extra-mural programs and mass communication.

The size of the training sector in development is today an important indicator of the seriousness of the development effort in a country and of the possibility of its success. The effectiveness of the development training sector and its efficiency are thus important considerations.

The Promise of Evaluation

The promise of evaluation in development training can now be well understood. Effective training of various trainers in the long trainer-trainee chain, and especially of the change agent at the front-line, is crucial. The size and scope of development work is such that good training can make an important difference to what ultimately happens to our schemes regarding food production, health and nutrition, family planning and political participation.

But to know where improvements in training are possible we need to undertake systematic evaluations of development training. Here is a set of questions that are crying for answers:

1. Had the new change agent roles designed to promote development within communities proved effective?

2. Had the institutions and programs established to handle training of new change agents performed their tasks effectively and efficiently?

3. Had the procedures for recruitment of change agents for development roles, the training curricula designed for their training, and instructional methodologies used in their training been correct?, and, finally.

4. Had the overall training effort made any significant impact on the lives of people whom the trained change agents were supposed to serve?

Answers to these questions will bring high returns on the investment in development training. Herein lies the promise of evaluation.

Summary

The development training enterprise is significant both in size and scope in most development countries. Improvement in the effectiveness of training can generate high dividends. To know how effective current training programs are, they should be evaluated. The promise of evaluation in development training is immense.

Things to Do or Think About

1. On the basis of your personal experience in development, what groups of development workers in your country you think is performing at less than a satisfactory level? Can more effective training help in that case?

2. In your opinion, which is the weakest link in the trainer-trainee chain within the development training sector in your country?

CHAPTER IV

DESIGNING TRAINING SYSTEMS FOR DEVELOPMENT WORKERS AND CHANGE AGENTS

To evaluate a thing, an instrumentality, a process or a system, we must first understand that thing, instrumentality, process or system. The first task that we face as evaluators of development training is to obtain an understanding of what development training is.

After understanding comes description. We must be able to describe the development training process. Since our ultimate objective in evaluating development training is to intervene in the training process to improve it, our description of the development training process must be made in design terms.

Education and Training

Distinctions have sometimes been made between education and training. While such distinctions may sometimes be quite useful, they should not be considered absolute. Both education and training, after all, are concerned with human learning and performance. Both seek to teach new values, new attitudes, new knowledge and skills. In this respect, education and training are similar. However, the two are dissimilar in one important way. Education is preparation for life in a society. Therefore, objectives of education are general. Training, on the other hand, is preparation for work in a particular structure or system. Therefore, objectives of training are much more particular. These distinctions are, by no means, sacred. Training, at its best, becomes education.

Training is, typically, defined in reference to some organizational objectives: a planned process of teaching and learning that seeks to prepare members of a group or an organization to acquire knowledge, attitudes and skills required by the group or the organization for implementing its objectives and, thereby, its overall mission.

Training Design: Theory and Tradition

Training design theory can be seen to be rooted in curriculum development theory. More recently training design has become directly linked with instructional development theory. Training design as practice has had a long tradition in the U.S. Army and in American business and industry. That is where most of its operational models and methodologies seem to have been developed.

The following six principles have emerged from research and practice on training within the U.S. Army over the years during and since World War II, as summarized by Bushnell:[1]

1. Performance-Based Instruction. Students should learn the skills necessary for job performance. The emphasis should be on active skill practice -- "doing" rather than "passive" adsorption of information.

2. Task Mastery. Every student should be required to reach a particular standard of performance in each skill. Assessment should be on a "go/no-go" basis. The student who does not reach the criterion level of performance should receive additional practice until he does reach it.

3. Functional Context. The student learns best in a job-relevant situation. Theoretical or technical material should be presented only when it is needed in learning to perform a skill.

4. Individualization. For various reasons people learn at different rates. To the extent possible, the trainee should be permitted to learn a skill at his own rate.

---

[1] Davis S. Bushnell, "Training as a Knowledge Production and Utilization Strategy: Instructional Technology in the U.S. Army," Viewpoints in Teaching and Learning, 54(2):114-127, April, 1978.

5. Feedback. If the instructor and curriculum developer know a good deal about the student to be taught, instructional methods can be modified to be more effective. If the student knows about his own skill acquisition, he will be motivated to correct errors of inadequate performance.

6. Quality Control. To ascertain that the training system is functioning properly, student performance must be systematically assessed at various times during and at the end of training.

Training within the army has been heavy on technology -- films, television, video, and computorized simulations.

In the business sector and in the few universities that are getting interested in the training technology, the new vocabulary is of front-end analysis, performance discrepancy, learner analysis, cognitive styles, task analysis, criterion-referenced testing, mastery learning, job engineering, environment analysis, contingency management, feedback systems and organization development.

Results obtained from the application of training design models and techniques listed above have been impressive within the formal organizational settings of the army and industry. Indeed, concepts such as front-end analysis, learner analysis, task analysis, individualization, feedback and criterion-referenced testing could all be usefully employed in the sector of development training -- training of development workers and change agents. However, one will have to be extremely cautious about a wholesale transfer of training models and techniques from the army and industry to socio-economic and political development within Third World communities. In the development training sector, the environment of work is fluid, and work roles have to undergo a process of adaptation which makes job engineering quite difficult. Learning by doing as a training strategy is not always possible within community development settings and role performers have to grow on the job. In this sector which has always been subjected to severe scarcity of resources, use of audio-visual media and simulation technologies has also been rare.

The Special Characteristics of Development Training

There are some special characteristics of training in the development work in general over the Third World. Some of these are discussed below:

1. Unlike the army or industrial training, training for development work has to deal with systems that are open and somewhat undefined. Development workers have to be taught to be sensitive to social systems; they have to learn to be culturally-oriented.
2. Development training also has to involve a component of needs assessment and needs negotiation processes. Development workers can not take pre-packaged solutions to pre-determined problems. In each new community, the development worker must go through the process of determining the "felt needs" of people and then "fashion" new profiles of needs, participatively with the local communities, through a process of needs negotiation, involving a dialectic between needs as seen by local communities and needs as defined by the developmental elite.
3. It follows from the above that training for development has to include both substantive and process components. Substantive knowledge in agriculture, family planning, health or intermediate technology is not enough, the development worker must have been trained also in communication and social participation skills.
4. Development training often involves the three-fold process of role invention, role capacitation and role socialization. Development workers often have to face the challenge of inventing new roles and introducing them in communities as they assist in capacitating and socializing

workers for those roles.

5. Typically, training is defined as preparation for role performance within a structure or an institution. However, in the development world, working in isolation within one organization or structure will be a disability. Development training has to involve an understanding of system interfaces and an ability for coordination.

6. Development work is in a very real sense political work. Invariably development involves new distributions of power, status and economic goods -- to include the marginal and the underpriviledged. To promote the welfare of the oppressed and the disadvantaged, especially within economies of scarcities, requires strong ideological commitment on the part of workers themselves. Thus development training must involve a component of value clarification and indoctrination.

7. While training always lends some legitimacy and status to the trainee, in the area of development training the allocation of status can have significant structural consequences within the community. The recruitment patterns used in the choice of trainees may perpetuate existing power structures or may introduce a new generation of non-traditional leadership from new social classes.

8. While effects of and returns from training within industrial and business settings could perhaps be measured in terms of dollars or shillings or some other currency, returns from development training are delayed and are often ambiguous. Development trainers have to develop some intermediate criteria of success for their training.

9. Finally, the realities of the world of development and of the Third World rural areas in particular, make control, supervision, follow up and

feedback quite difficult, if not impossible. Recruitment of trainees itself can not always be controlled, they may be self-selected. Nor is it always possible to control trainee performance. It is not possible to integrate voluntary development workers within organizations and to supervise their work. Lack of infrastructures of communication often make follow up and feedback well-nigh impossible.

A Model of Development Training Design

Training design for development settings requires "system thinking" on our part. What is a system? And what is "system thinking"?

A system is an orderly arrangement or combination of interrelated and interdependent parts or elements emerging into a whole. A family is a system. A cooperative is a social system. A training institute is a social system -- maybe we should say a techno-social system.

System thinking is the mental habit of looking at things, whole. It is "holistic thinking." It is the type of thinking that enables us to avoid getting caught into one-to-one relationships, and linear extensions. It is thinking that involves "atonceness" and learning to see systems where others may see mere individual entities.

System theorists have studied the properties and characteristics of systems. They have found that all systems can be described in terms of four parameters (or guidelines): inputs, processes, contexts, and outputs. That is, all systems receive some inputs which are subjected to some processes, under particular contexts, to produce some outputs. The ability to use the same four parameters to describe all systems is most convenient and a most useful consideration.

A generic model of training design in developmental settings is presented on the following page. It uses the system metaphor and the four system parameters -- inputs, processes, contexts and outputs. This model should help trainers in the developmental sector to organize all the variables that can or do enter a training situation. They can then look at those variables that offer real options for manipulation within a particular training setting. If options are available, they can then evaluate variables under different conditions and put evaluation to work to actually improve the performance of training systems.

The model presented here shows relationships between and among three systems -- the training system, the performance system, and the social system. The outcomes of the training system, with additional inputs XYZ (landrovers, medicines, high-yielding variety seeds, etc.) become inputs for the performance system. In turn, the outputs of the performance system with additional inputs PQR (ideologies, trained planners, infrastructures, etc.) become inputs in the larger societal system engaged in development.

A systemic view of development training, with mutual interfaces between the training system and the program system, and then between the program system and the social system, should enable training designers to ask questions about the development roles that should be designed, and about the technical and social content of those roles.

Steps in Training Design

In the following, we have translated the general model of development training design into "action terms" and have suggested six steps involved

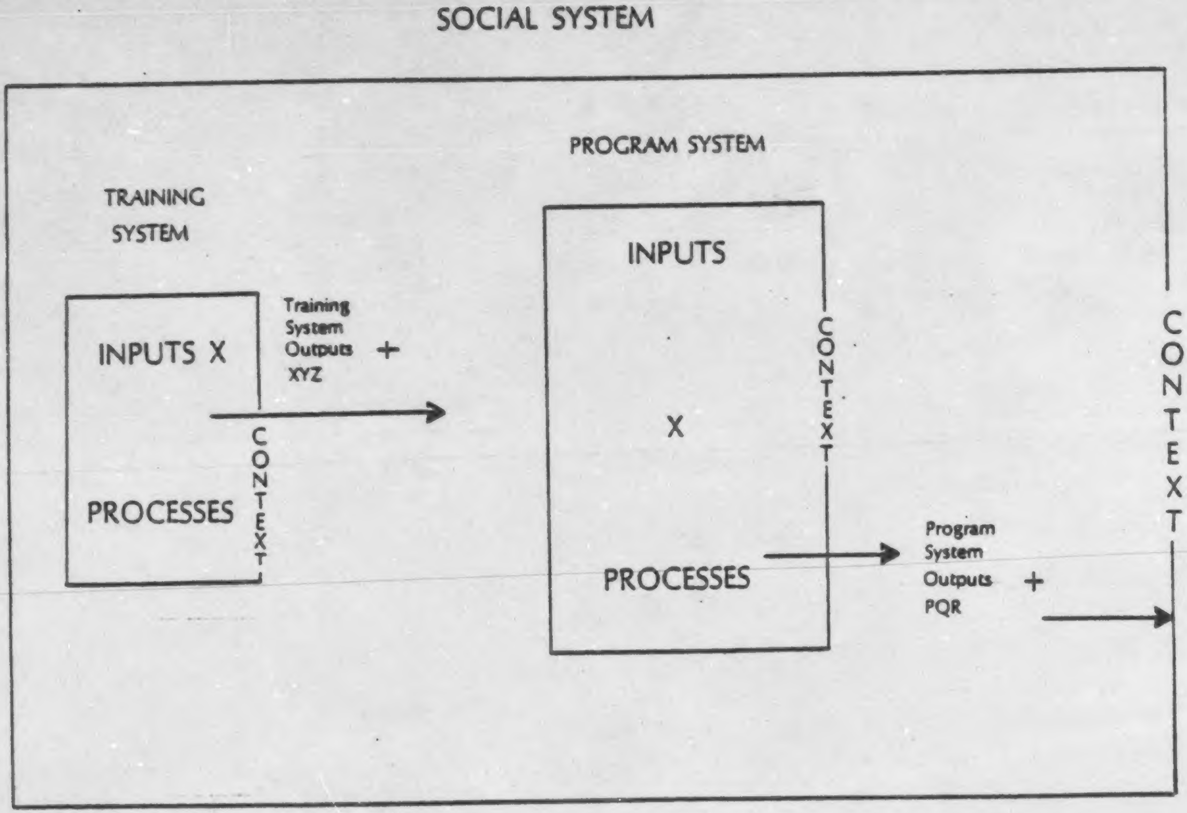

SOCIAL SYSTEM
PROGRAM SYSTEM
TRAINING SYSTEM
INPUTS X
PROCESSES
CONTEXT
Training System Outputs XYZ
+
INPUTS
X
PROCESSES
CONTEXT
Program System Outputs PQR
+
CONTEXT

Figure 1. A model of development training design and evaluation. (Bhola, 1980)

in training design:

STEPS IN TRAINING DESIGN[1]

Step I	Social and program analysis for the definition of both general and specific development objectives
Step II	Listing program activities to be conducted for the achievement of defined development objectives
Step III	Defining and designing roles and allocating role responsibilities for conducting listed program activities
Step IV	Statement in behavioral terms of competencies required of various role incumbents; and listing of new attitudes and values needed for socialization in those roles
Step V	Developing a training curriculum to build required competencies and the implementation of this curriculum in training courses, workshops, on-the-job training and correspondence training
Step VI	Evaluation of the effectiveness of training and review of steps I to V.

One should note that the steps for training design listed above encompass all the three processes of role design, training for the role and socialization for the role.

By way of a summary, the concept of development training design can be defined as follows:

---

[1] Adapted from H. S. Bhola, "Description and Evaluation of a Training Programme for Literacy Teachers and Their Trainers," Indian Journal of Adult Education, Vol. 31, No. 5, May 1970, pp. 3-6, 14-16.

- To train is to instruct and socialize
- in relation to a defined set of organizational objectives,
- using appropriate content and methods,
- in appropriate settings;
- within a temporary system or an established institutional structure,
- an appropriately selected group of functionaries, generally from a larger pool of such functionaries, using some particular criteria;
- for effective role performance,
- at some anticipated level,
- within a matrix of organizational and social realities of the world of work; and
- with expectations of feedback for trainers and of self-development by trainees on the job.

Havelock and Havelock[1] list four steps in training design as follows:

1. Planning
2. Defining objectives
3. Specifying learning that should meet objectives, and
4. Specifying the sequence of training activities that should lead to desired learning.

They then go on to list eight principles of good training design:

1. Relevance to personal background and back-home situation of trainees
2. Specificity of objectives
3. Generality so as trainees will be able to cover a broad range of work

---

[1] Ronald G. Havelock and Marcy C. Havelock, Training for Change Agents: A Guide to the Design of Training Programs in Education and Other Fields. Ann Arbor, Michigan: Center for Research on Utilization of Scientific Knowledge, University of Michigan, 1973.

situations and problems

4. Reinforcement and rewards for appropriate responses and actions of trainees
5. In-process evaluation and feedback
6. Cost-effectiveness of training
7. Synergy (use of a variety of media and methods), and
8. Transferability of learning and skills from the training situation to the work situations.

Another training specialist, Odiorne[1] focuses on the training process and provides suggestions for "instructional development" in training settings:

1. Specify the desired terminal behavior or the exit requirements of the trainee.
2. Specify the present level of behavior and performance of the trainee.
3. Using (1) and (2) above, define clearly the specific behavioral change required in the trainee.
4. Provide situations within which the trainee can engage in active action such as talking, conferring, writing, operating, demonstrating.
5. In implementing (4) above simulate behavior sought on the job as closely as possible through role plays, case studies, management games, and demonstrations.
6. Guide the course of change of behavior through an orderly progression of small steps.

---

[1] George S. Odiorne, Training by Objectives. New York: The McMillan Co., 1970.

7. Provide feedback through knowledge of results at each possible stage of training.

8. As the course instructor, stay in control of the learning procedure.

9. Measure actual outcomes through evaluation.

As can be seen, the three sets of instructions listed above complement each other. Bhola's suggestions and steps raise the design questions within the larger framework of development programs and training institutions. The process of "role design" is considered by Bhola to be a significant part of the training design process in development settings. Havelock and Havelock focus on the training system, paying due attention to the world of performance where the trainees will go to work. Odiorne discussed training design as an instructional development problem. Each of these three aspects must, of course, get due attention in an overall training design effort.

It is the task of the training designer to bring together, in different values, the many variables of a training system into an ideal mix to create the best training plan under a given set of circumstances. The training plan should be such that:

1. it is ideologically and strategically congruent with the national vision and the national development strategy.

2. it uses existing national institutional resources for training where such resources exist, and establishes new institutional mechanisms where such mechanisms are needed.

3. it chooses training objectives, methodologies and durations for training that the society can afford in terms of resources that will have to be used.

4. it relates directly with the cultural and socio-political realities of the setting in which trainees will perform after training.

5. it is based on an understanding of the learners and communities to be served in regard to their values, motivations and their own perceptions of needs and priorities among such needs.

6. it takes in view the existence of other change agents and institutions concurrently working in the field and with whom coalitions and collaborations may have to be built.

7. it seeks to teach trainees both substantive knowledge and process skills using appropriate materials and methodologies in well-chosen training settings; and

8. it maximizes outputs of the training system in terms of trained development workers, experienced trainers, and tested training strategies and training research; and minimizes the total array of training costs. (See the "Training System Analyzer" included as Table 6.2 in the monograph, Curriculum Development for Functional Literacy and Nonformal Education by H. S. Bhola, 1979, pages 141-148.)

Some Training Models in Use

In the practice of training design in development settings, some special models for the delivery of training to development workers and change agents have emerged over the years. We will here refer briefly to three of these models:

1. The Operational Training Model
2. The Participatory Training Model, and
3. The Action Training Model.

1. The operational training model

The operational seminar[1] approach was developed within Unesco over a period of many years in the context of literacy promotion for development.

In the words of de Clerck:

> "The operational seminar may be described as a training exercise which enables the participants to experience, on a reduced time-scale but in a real situation, the sequence and correlation of operation, which taken together, constitute the process of literacy work linked to development, whether the development be predominantly economic, social or cultural. Like research workers operating in problem situations, the participants rediscover for themselves, through an individual and group effort of research and analysis, an educational strategy geared to a development strategy and the fundamental concepts of a new type of literacy work. They acquire direct experience of ways of preparing and shaping a 'mode-to-measure' programme rooted in the requirements of a specific milieu, and a practical knowledge of experimenting with educational activities and assessing the various factors in the learning process. ... Theories and concepts are constantly held up for comparison with the practical realities of a particular environment."

The operational seminar methodology thus seeks to be the methodology of "method demonstration" or "demonstration by result." The participants, through their involvement in a real-life task, are supposed to become instrumental in their own training.

Since 1970, more than 50 operational seminars have been organized all over the world, covering some 3,000 trainees. Participants have included all levels of cadres, from planners to front-line workers, agricultural technicians, health technicians, adult educators and social scientists -- all in the same one seminar group. The seminars typically last 3 weeks though some have been shorter, and many have been longer. Participants work in multidisciplinary teams within a particular community or zone of activity. The level of direction of their activities depends upon their needs. A final report of the work of multidisciplinary teams is produced at the end of the seminar.

---

[1] Marcel de Clerck, The Operational Seminar: A Pioneering Method of Training for Development. Paris: Unesco, 1976.

While the model is well conceptualized and has undergone a long period of testing, it has suffered from some serious problems in terms of professional guidance of participants; in regard to entry and exit from communities; relationship between theory and practice and, therefore, in regard to transfer of learning from the training situation to the work situation; and about the quality of learning as demonstrated by the reports produced by groups.

2. The participatory training model

Paulo Freire, the Brazilian philosopher, educator and author of Pedagogy of the Oppressed (Herder and Herder, New York, 1972), is the man who scratched the word "participation" indelibly on the conscience of development planners, trainers and evaluators. All the world over, groups of people have worked with Paulo Freire's ideas and extended and enriched them through implementation.

Kamla Bhasin's work in the area of participatory training also builds upon Freire's ideas and has the merit of being well-documented.[1]

The methodology of participatory training is non-directive, dialogical, and experiential. The following, in her own words, are the main features of her approach:

---

[1] Kamla Bhasin, Participatory Training for Development. Bangkok, Thailand: FAO, 1976. Also by Kamla Bhasin, Breaking Barriers: A South Asian Experience of Training for Participatory Development. Bangkok, Thailand: FAO, 1979.

1. It was attempted to involve the participants in identifying their own training needs, in planning and in running the programme because the best way to teach about 'bottom-up planning', 'people's participation', and 'decentralisation of authority' is by practicing these ideas.

2. The emphasis was on self-training and group learning through continuous interaction between participants and on exchange of experiences through group-discussion and actual field visits. This was done on the assumption that by exposing the participants to the realities of rural development in their respective projects, they would be able to

- identify the issues and problems that confront them as change agents;
- arrive at solutions which are most applicable to their own situations; and
- act these solutions within the framework of their own organizations and the communities in which they work.

3. All learning during the programme was related to the concrete problems of rural areas and their possible solutions. The raw material of learning was provided by the participants' actual experiences in the course of their work. This was achieved through case study presentations, field visits and analytical discussions.

4. The emphasis was on group-discussion rather than on lectures.

5. Attempts were made to create an atmosphere in which the participants would feel free to express themselves and be ready for frank reflections and analysis. In other words be prepared for an honest, even if painful process of self-searching, criticism and self-criticism.

6. It was attempted to run the programme itself into a process of living together in a collective spirit.

7. An ongoing evaluation was built into the programme.

The main components of the training methodology were the following:

- Participation of trainees in planning and running the training
- Collective living
- Case study writing and presentation
- Field visits
- Group discussions.

3. The action training model

The action training model (ATM) emerged within the context of a series of training workshops and panels on the evaluation of development training programs conducted under the aegis of the German Foundation for International Development in Kenya, during 1979-1982.

The following are the important features of the ATM:

1. A long period of pre-planning and preparation. An important feature of the ATM is the long period of pre-planning and thorough preparation for the planned workshop or seminar. The organizers prepare themselves not for one particular pre-packaged workshop, but for an unanticipated workshop, any one workshop out of a series of possible workshops. Such preparation makes it possible for the workshops to be participatively planned.

2. Participative planning of workshops. The workshops are planned participatively in terms of the needs of the group at a particular historical time in the life of their development program. The workshop for which the organizers had prepared themselves is locally re-invented by the participants.

3. Participants control content, emphasis and schedule. The content of teaching-learning, the emphasis given to various topics and the schedule of activities are controlled by the participants. The workshop works according to a rolling plan -- everyday, the program for the rest of the workshop is replanned. Schedule for the day is printed at the end of the day.

4. A learning community. The participants became members of a learning community who live, work and play together. Everybody lives on the same location using the same level of facilities. Late arrivals and early departures are not allowed.

5. Facilitators, not teachers. The organizers look at themselves as facilitators rather than teachers. Contributions are encouraged from

everyone. Teaching-learning methods include lectures, discussions, panels, group work, and individual tutoring, depending upon the needs.

6. The action orientation. Emphasis is on relating theory to practice; to enable participants to do something practical with their ideas. During the workshop, each participant is asked to prepare a proposal for an evaluation study. This evaluation study must be implemented during the year between the workshop they are then attending and the next workshop at which they must report on the completed study.

7. Continuous evaluation and feedback. The workshops are continuously evaluated by using formal instruments as well as through the daily steering committee meetings every night where drop-ins are not only allowed but encouraged.

8. The integrative role of documentation. Special documentation is prepared for each workshop. First, this documentation assures that no lecture notes need to be prepared and distributed at the workshop. All the time available can be spent on learning. Also, it provides the trainees with the big picture of the area of evaluation to ensure that participants do not confuse part with the whole. What they may have decided together to study during the workshop may not be the whole of the evaluation process and of the evaluation knowledge.

9. Institutionalization for local transfer of responsibility. There is emphasis on the capacitation of local professional personnel and on the building of institutional capacity for the transfer of responsibility. Local faculties are associated with the organization of workshops and given time and provided resources to assume leadership. An Educational Evaluation Resources Committee, for example, was established in Kenya as a

step towards the institutionalization of the evaluation capacity in the country. During 1982-86, the host institution for the DSE workshops will indeed be on its own in offering evaluation workshops to trainers of development workers and change agents.

Developing Descriptions of Training Programs in Design Terms

In the opening paragraph of this chapter, we had suggested that evaluators must be able to understand what they seek to evaluate, and must be able to develop adequate descriptions of the phenomena to be evaluated. Since the purpose of evaluating training programs, typically, will be to intervene in the training process to improve it, these descriptions of development training programs must be in design terms.

To develop such descriptions we must once again engage in system thinking and return to the model of development training design and evaluation included earlier in this chapter. This time we should use the four parameters of systems (inputs, processes, contexts and outputs) to list the variables involved within the universe of development training.

The chart on the following page provides a first list of variables that will typically appear under the four parameters (inputs, processes, contexts, outputs) in the case of each of the three systems (training system, performance system, and social system) that appear in the model referred to above. Can you think of any other variables?

INPUTS	PROCESSES	CONTEXTS	OUTPUTS	
A. Trainers B. Trainees C. Materials and facilities	A. Instructional/ formative B. Organizational/ structural C. Distributive/ maintenance-related	A. Organizational B. Environmental (extra-organizational)	A. Trained development workers B. Experienced trainers C. Training research and training materials D. Unanticipated results	TRAINING SYSTEM
A. Trained field workers B. Learners in communities C. Learning materials and facilities D. Economic inputs E. Institutional inputs	A. Planning and needs assessment B. Organizational/ coordination C. Animation/ motivational D. Instructional/ socialization E. Distributive	A. Organizational B. Communal C. Ethnic/ linguistic D. Rural/urban E. Socio-economic	A. Informed and resocialized youth and adults B. New institutions C. Developed and self-reliant communities D. Unanticipated results	PERFORMANCE SYSTEM
A. National mission B. Leadership/ manpower C. Informed citizenry D. Institutions E. Technology/ material resources	A. Ideological/ policy making/ planning B. Legislative/ organizational C. Communication/ mobilizational D. Infrastructural/ distributive	A. International/ world system B. Cultural C. Political/ socio-economic	A. A developed and just society B. Unanticipated results	SOCIO-POLITICAL SYSTEM

Figure 2. Variables under each of the four system parameters of the training, performance, and social systems.

The trainers' universe and the trainers' tasks

The generic model of development training design presented earlier, together with the chart of parameters and variables that has followed the model should enable us to chart the universe of training in the area of development. It should help us see the variables that will enter the training system and to see the variables of the performance system and of the social system that will directly influence training design and implementation.

In the next chapter we will deal with the issue of generating evaluation questions and discuss the criteria for choice among such questions to be able to develop useful evaluation agendas for evaluating the effectiveness of development training.

Summary

To be able to evaluate development training, we must understand the process of training design. While training design has a long tradition of theory and practice in the army and industry, the development training in the Third World settings raises some unique problems. Development training must often involve more than capacitation for the role; it must also involve "role invention" and "role socialization." A model for the design of development training programs is presented which can also be used for adequate descriptions of ongoing training programs.

Things to Do or Think About

1. Sketch the main features of the training model or approach in use in your training center or program (as we have done in this chapter for the Unesco's Operational Seminars, the Participatory Training Model and the

Action Training Model).

2. Using the figure on page 51, analyze the input-process-context-output variable in the case of your training program.

CHAPTER V

ASKING EVALUATION QUESTIONS, MAKING EVALUATION AGENDAS, AND TYPICAL EVALUATION THEMES

There will be situations when the trainer-evaluater will have an evaluation question given to him or her. Ideally, however, it should be possible for the trainer-evaluater to go through the systematic process of describing the total system of training; be aware of the options typically available under each of the input, process, context and output variables operative within the training system; on the basis of problems experienced or the potential visualized, generate a set of evaluation questions; and choose from among these evaluation questions, those that must be answered on a priority basis.

The First Step: Description

In Chapter IV, we have already presented our approach to developing descriptions of training systems. A model of development training design and evaluation was offered, followed by a chart listing all the variables under the four system parameters (inputs, processes, contexts and outputs) for each of the three systems (training system, performance system and socio-political system) in interface with each other.

An outline chart as shown on the next page should be used to develop a description of the training system in which the trainer-evaluator is involved. More attention should be paid to the row dealing with the "training system." The other two rows dealing with the "performance system" and the "socio-political system" could receive less exhaustive attention. The entries in the cells of the chart should not be general

INPUTS	PROCESSES	CONTEXTS	OUTPUTS	
				TRAINING SYSTEM
				PERFORMANCE SYSTEM
				SOCIO-POLITICAL SYSTEM

Figure 1. An outline chart for use in developing a description of the training system in which the trainer-evaluator is involved.

but should include specific and concrete statements of the realities within the training system being described.

Step Two: Know What the Possible Options Are

Trainers-evaluators should not only know what specific variables enter into the concrete situation of their training program but also know what options are typically available under these variables. As students of training design they should know "what is," as well as "what can be." While it is often possible for trainers to influence the choice of options within the performance system and maybe even in the larger socio-political system, in the following table we list and discuss options available within the training system only:

PARAMETERS / Variables		Options
INPUTS		
	Trainers	Educational levels and knowledge of subject; their social class and value orientations
		Extent of field work experience
		Level of commitment to development work
		Teaching competence and teaching experience
		Direct appointment versus secondment from a parent department
		Continuity versus turnover
		Work loads of trainers
		. . . .
		. . . .
		. . . .

Trainees	Educational background
	Pre-service versus in-service groups
	Social class and value orientations
	Commitment to development work
	Motivation to learn
	Recruitment modes (deputed, sponsored, open competition)
	. . . .
	. . . .
	. . . .
Teaching Materials and Facilities	Teaching materials--quantity, diversity, quality
	Indigenous versus imported instructional materials
	Instructional and duplication equipment
	Physical plant of the training center
	Transportation for field visits
	Budget
	Level of program commitments
	Access to experimental villages and pilot projects, etc.
	. . . .
	. . . .
	. . . .

PROCESSES

Instructional/ formational	Conceptualization of training as knowledge transfer, skills training, behavior modification, socialization, etc.
	Integrated versus discipline-oriented curriculum development (i.e., instructional organization)
	Task-specific versus training for core skills

	Teaching and learning styles
	Substantive knowledge versus process emphasis
	Presence versus absence of curriculum validation through needs assessment
	Conceptual versus field operational orientation
	Availability or nonavailability of counseling and guidance services
	Instructional delivery -- face-to-face, distance, etc.
	. . . .
	. . . .
	. . . .
Organizational/ structural	Organizational health status
	Organizational capacity rating
	. . . .
	. . . .
	. . . .
Distributive/ maintenance-related	Quality of administrative support
	Residential facilities for staff and faculty and for trainees
	. . . .
	. . . .
	. . . .
CONTEXTS	
Organizational	Organizational culture
	Institutional relationships (horizontal and vertical) with other organizations
	. . . .
	. . . .
	. . . .

Environmental	Surroundings (Closeness to a bar versus a "retreat" situation)
	General social climate in the country
	. . . .
	. . . .
	. . . .

OUTPUTS

Trained development workers

Emergent role identities

. . . .

. . . .

Experienced trainers

. . . .

. . . .

Training manuals

Radio programs

Correspondence courses

. . . .

. . . .

New professional stratifications

Contributions to a "brain drain"

. . . .

. . . .

Step Three: Generating Evaluation Questions

The trainer-evaluator must now confront "What is" with "What can be." The trainer-evaluator should now look back critically on his or her day-to-day experiences within the training program and try to articulate clearly the problems which were there but perhaps were hard to get hold of. The trainer-evaluator should also look at the existing training system positively and think of the higher returns that could be obtained by making some changes.

In all these cases, the trainer-evaluator should be able to state some information need: We have the problem "X", but we do not have the information "Y". Or, if we had the information "Y", we could take the promising step "X", with confidence.

Two important points must be mentioned here.

1. A distinction should be made between evaluation problems and administrative problems. To administer is to direct and superintend the execution, or conduct of a program. If administrators, for reasons of incompetence or for lack of responsibility fail to direct and superintend a program, the problem is administrative, not that of evaluation. Evaluation can only assist administration by providing needed feedback data and by testing various program assumptions. It is not a substitute for good administration.

2. Evaluation may require analytical evaluation or collection of framework data. There will be evaluation questions which will have to be answered through conceptual and operational analysis rather than by going to the field for collecting data. Is the participatory method claimed to be used in a training program, actually employed in the training protocols? Is the integrated curriculum concept actually embedded into the training plans, training materials, and training delivery and schedules? These questions require analytical answers and not necessarily collection of data. Again, policy documents, the nation's five-year economic plans and census data may have to be used for building a framework to the evaluation of an aspect of the development training program.

Step Four: From Evaluation Questions to Evaluation Agendas

The evaluation questions generated in Step Three above may all be interesting and promising but it may not be possible to answer all of them in the particular context of a training program and within the resources available. In such a situation, a particular evaluation agenda must be made to be followed within a particular time period.

The following criteria might be useful in the choice of evaluation questions for inclusion in the evaluation agenda:

1. Availability of design options
2. Significance of the evaluation question
3. Feasibility of implementing the evaluation study.

(1) Availability of options. All of the variables entering a training situation may not be under the control of the training specialist. In other words, the trainer may not be able to change the values of the variables in any significant way. If such is the case and training variables are immutable, it is no use evaluating them because they do not offer design options anyway.

(2) Significance of the evaluation question. If a variable does offer a design option, it will make sense to evaluate it, if in a relative sense, it offers a significant option. The significance has to be in terms of results in the effectiveness of training or in its efficiency. In either case, the returns from the evaluation effort should be worth the effort.

(3) Feasibility in regard to available resources. The evaluation question chosen and the evaluation design that is necessary for conducting the

evaluation should be within the capacity of the training center or institution. Reasonable amount of resources should be available for evaluation to avoid unnecessary frustrations.

Typical Evaluation Themes in Development Training

The central function of evaluation of training is obviously to validate the assumptions on which training design has been based. To put it differently, trainers engage in evaluation to judge the worthwhileness of their training programs, in terms of some criteria, on the basis of information that can be collected or is available.

The basic aims of the evaluation of development training can be stated as follows:

1. measuring the abilities, skills and value orientations acquired by trainees during the various stages of a training cycle or at the end of training; and, later, to measure performance on the job in actual field situations.
2. on the basis of the above, providing feedback to trainers in regard to choice of training objectives, content, methods, materials and instructional settings for appropriate remedial actions, and
3. providing information to policy makers and planners in regard to the overall training policies, designs, patterns of recruitment, role definitions, performance in the field by role incumbents, institutional settings of training and the total allocation of resources to the training effort.

These basic aims of the evaluation of development training will typically appear as the following evaluation themes:

1. Base-line survey
2. Needs assessment
3. Curriculum evaluation
4. Achievement and attitude testing
5. Institutional or organizational evaluation
6. Cost-effectiveness evaluation
7. Evaluation of impact
8. Management information system, and
9. Monitoring and quick appraisal.

1. Base-line survey

Base-line surveys of communities are undertaken to establish the economic, social and cultural base-line against which later changes can be measured. Community development workers generally would conduct extensive base-line surveys in communities they seek to serve. Wherever possible, trainers-evaluators should use already available base-line data to design their training programs for development workers and change agents. It is possible, however, that the base-line survey already conducted had not anticipated the special information needs of trainers-evaluators.

The special information needs of trainers-evaluators may deal with (1) role considerations, and (2) knowledge considerations. A trainer-evaluator preparing family health education workers, for example, would need to know the current child-rearing and health practices within communities, level of knowledge of nutrition, lack or otherwise of home gardening; and level of consumption of animal proteins. At the same time, the trainer-evaluator would be interested in how this knowledge is currently acquired by mothers; whether traditional educational roles exist that disseminate this information; what other more modern-secular roles have already been introduced within those communities by the government; and what expectations one should have about the introduction of a new role of the family health education worker.

As can be surmised, trainers in cooperative training colleges or trainers of agricultural extension workers will have to design base-line surveys to fit their special information needs about existing economic knowledge, behaviors and structures; and the design of roles of extension agents.

2. Needs assessment

Trainers-evaluators may have to conduct needs analyses at various systemic levels. They may conduct a general needs analysis at the national level to project those needs in training design. They may also do a needs analysis at the performance system level to see what demands a particular program or project will make on development workers and change agents. Finally, needs assessment may be conducted within the training system itself to design appropriate knowledge, value and performance content of a training course or workshop.

A good needs assessment, typically, will cover all the constituencies involved within a training system -- trainees, their field supervisors, development administrators, community leaders, and communities themselves. The final training design should be done on the basis of the various needs profiles generated by these different constituencies and groups.

3. Curriculum evaluation

Evaluation of training curriculum will be a frequent evaluation theme of trainers-evaluators. They may need to evaluate particular items of instructional materials -- a primer, a handbook, a set of charts, a simulation-game. They may want to evaluate a particular training method, for example, team facilitation versus single tutor. Different systems of program delivery may be tested: correspondence courses versus night schools; teaching mothers or teaching families, etc. Finally, the overall effectiveness of a training curriculum may be the concern of trainers-evaluators.

4. Achievement and attitude testing

A considerable part of evaluation within a training setting will consist of testing. It will be testing of trainees as they enter the training system and their testing as they leave. This testing will cover knowledge; diagnostic and performance skills; motivations, attitudes and values; and communication and process skills.

Some of this testing may have to be done not on the trainees but on individuals in the communities, to be able to judge change agents' performance by the impact they may have had in the communities.

5. Institutional or organizational evaluation

The quality of institutions or organizations determines the quality of services these organizations will be able to produce and deliver. Unforunately, very little attention seems to have been paid by development trainers (or by development specialists generally) to organizational traits. Institutions or organizations can be studied along two general dimensions: (1) organizational climate, and (2) organizational capacity. Organizational capacity is determined through an accounting of an organization's resources in relation to its mission. Organizational climate is a conceptualization of an organization's social life -- member's identification with the organization and their satisfaction or dissatisfaction with organization's decision making style and patterns.

6. Cost-effectiveness evaluation

Two terms are in use in the literature of evaluation in the development sector: cost-benefit analysis and cost-effectiveness analysis. Both these analyses involve comparisons of costs and outcomes, but the nature of comparisons differs.

Cost-benefit analysis is possible when outcomes can be given clear economic values in dollars and cents. This sort of economic analysis is seldom possible in education and extension where non-material effects are the most significant but can not be assigned material values.

Cost-effectiveness analysis is used where outcomes can not be expressed in monetary terms because of the absence of market prices for outcomes.

Therefore, the levels of outcomes themselves are compared in proportion to the costs incurred in each different case.

It may be also useful for trainers-evaluators to work out unit costs, i.e., costs per trainee trained, to be able to allocate resources more effectively.

7. Evaluation of impact

The study of the impact of training on change agents must go beyond testing of trainees within the organizational setting. Trainers-evaluators must go to the communities where their trainees work. Their questions must, however, be sharply focussed: Did the change agent role fit into the social setting? Was the role performer able to teach, demonstrate and resocialize? These questions will be possible to answer if base-line data was collected earlier.

The study of the impact of change agent roles on communities must provide proper time for the new role to go through the period of adaptation and use by the communities. They should have time to relate, learn and adopt. Such "latency periods" may have to be many months (if not many years) long.

Also, in the study of the impact of new roles and new teaching within communities, trainers-evaluators should look for both the anticipated and the unanticipated consequences of the introduction of change agents. Have role conflicts emerged in relation to traditional roles? Is a new group of power holders emerging within communities because of new role performers? Has the change agent brought in bureaucratization; resulting in the destruction of local initiatives?

8. Management information system (MIS)

Trainers-evaluators can not depend on special evaluation studies for all their information needs. Special evaluation studies take time and resources, while decision-makers (planners, administrators, trainers) need information regularly to make day-to-day decisions. For fulfilling these daily needs of information, trainers-evaluators must build management information systems (MIS's). An MIS is developed by the systematic collection and storage of data that is generated by a program in the very process of its implementation. All that is required is that data recording devices such as registers, application forms, grade books, diaries, log books, supervision report forms be devised and their use be mandated. Supervision reports and periodical reports be regularly obtained and filed, as also instructional materials, and related records and documents.

Such data should be processed regularly at various decision points to make appropriate management decisions. It should also be used for secondary analysis as part of special evaluation studies. In other words, special evaluation studies should be able to use, whenever possible, the data already stored in the MIS.

9. Monitoring and quick appraisal

In Chapter I, we made a distinction between evaluation and monitoring. The dictionary meanings of monitoring are to watch, observe, check and sometimes adjust. Monitoring is thus a term used for collecting the status information from the operating and implementation levels of a program to determine performance in comparison to expectations. It should be obvious that a good MIS will help the monitoring process.

Another term coming in use in evaluation literature is "quick appraisal." Quick appraisal is the child of necessity. Systematic evaluation takes time, yet in many instances evaluation needs may be both urgent and compelling. A quick appraisal will typically be done by a visiting team of evaluators using small samples and short tests and schedules, with findings provided to decision makers within four to six weeks. These findings will be in the form of strengths and weaknesses in the program; factors that seem to be responsible for those strengths and weaknesses; and what corrective actions seem clearly called-for.

Summary

To ask good evaluation questions and to make good evaluation agendas, trainers-evaluators must engage in "system thinking" and begin with descriptions of training programs for which they are working. They should not confuse administrative problems with evaluation problems. In choosing evaluation questions for study, the criteria used should be (1) the availability of options to make changes in the training system on the basis of evaluation results; (2) the significance of the question in terms of returns on the time and resources committed to the evaluation study; and (3) the feasibility of conducting the evaluation study. Typical evaluation themes in which trainers-evaluators will be interested are base-line surveys, needs assessments, curriculum evaluation, achievement and attitude testing, institutional or organizational evaluation, cost-effectiveness evaluation, evaluation of impact, development of management information systems and monitoring and quick appraisals.

Things to Do or Think About

1. What are some of the options available in the context of your training program regarding recruitment of trainees, teaching methods and materials, or location and delivery of training? What are some of the information needs for you to make sensible choices among available options? State these information needs in the form of evaluation questions.

2. What do you need to know about the most significant demands made upon development workers and change agents by the "performance system" in your case? Reflect this need for information and knowledge in your evaluation questions.

3. What are some of the important facts of socio-economic and political life in your environment that must be reflected first in training design and then in the statement of evaluation questions?

CHAPTER VI

WRITING FORMAL PROPOSALS FOR EVALUATION STUDIES

Successful, cost-effective and timely completion of an evaluation study requires considerable forethought and pre-planning. This thinking and pre-planning can be best done within the framework of developing a "formal" proposal for the evaluation study. The process of developing a proposal for the evaluation study can be used to systematize the evaluator's own thinking; to clarify technical, secretarial and material needs of the study; to take stock of available resources; to request and receive consultant help, if necessary, on various aspects of the evaluation study; and to use the proposal as a tool of communication with administrators and interested parties.

Proposals for evaluation studies dealing with the effectiveness of development training in the various development sectors, we suggest, should include the following elements:

1. The developmental context of the training program.
2. The description of the training program in design terms.
3. The larger "complex of problems" within the total training program.
4. The evaluation problem chosen for study.
5. Justification for the choice of the evaluation problem chosen for study.
6. A review of available knowledge directly or indirectly related to the evaluation problem chosen for study.
7. The questions and sub-questions to be answered by the evaluation study.

8. The evaluation model(s) or approach(es) chosen for the evaluation study.
9. Evaluation design (or steps and procedures) to be used in conducting the evaluation study.
10. Instruments and tools to be used for data collection.
11. Plans for library and documentary research; and plans for field work where collection of field data is involved.
12. Plans for analysis of data and preparation of graphics, if needed.
13. Budgetary plans.
14. Reporting plans: who to report to, when, and in what form?
15. Bibliography, resource persons to be used, etc.

A beginner, writing his or her first proposal for an evaluation study, may find it useful to go through steps 1 to 15, more or less, in the order given above. The more experienced proposal writer may be able to jump back and forth to various steps: from step 4 to step 7, to step 10, to step 12 and so on. Again, in the settings of workshop and seminars of short durations, it may be necessary to focus on some steps and not on others.

It should also be kept in mind that until the final proposal is ready, the various parts of the proposal will require constant review and revision. The development of tools and instruments may require a look back at the indicators chosen for the study. A review of the indicators may require rewriting of the evaluation question and of its justification. Even after the proposal is all done, the realities of the field may demand changes and revisions, once again. One should be mentally ready for these never-ending reviews.

We will now elaborate and expand upon the various elements listed above, citing examples where necessary.

1. The developmental context

Development requires learning of new knowledge, new skills and new attitudes by workers, farmers and housewives. To deliver this learning to those who need it, we need trained change agents. The role and functions of the training institute or training program of interest to the evaluator should be put within the development context. The training program's contribution to the national effort in the training of manpower for development should be clearly brought out.

A statement of training objectives should follow. If the institution offers a variety of training programs, each different program should be listed, with specific objectives of each program indicated separately. In some cases, it may be useful to include the organizational chart of the training institution or program. (See Chapters II and III.)

2. The description of the training program in design terms

A description of the training program in design terms must be developed. First, the general characteristics of the training approach should be listed:

(a) Is the training supposed to be general or specialized?

(b) Does it emphasize teaching of critical consciousness or the dissemination of knowledge and skills?

(c) Is the training planned participatively or is it formal and prepackaged?

(d) Is the training offered academic or operational? and

(e) Does the training seek to teach entrepreneurial values or communal values?

There may be some other important questions that could be asked, but the above list should provide a good starting point.

These general questions about training design must be followed by a description of the training program to be evaluated in system terms. The four system parameters (inputs, processes, contexts and outputs) should be used to describe the training system in concrete terms and values. (See Chapters IV, especially the model of development training design and evaluation on page 39; and the chart including variables under each of the four system parameters of the training, performance and social systems on page 51.)

3. The problem complex

Evaluation problems arise from a lack of information or a lack of understanding. We may have no information or insufficient information on inputs and about the context of our work. We may have less than adequate understanding of the processes and their application within our particular setting. We may have no measure of the quantity or quality of our outputs. These shortcomings together will create a whole "complex of problems" in any training program. Indeed, a training institution or a training program is unlikely ever to be short of evaluation problems.

In developing a proposal for an evaluation study, an evaluator should review the whole set of interrelated problems found to be bothersome to program administrators and decision makers. The evaluator must, however, distinguish between evaluation problems and purely administrative problems.

Evaluation problems arise from lack of information and understanding, whereas administrative problems arise from incompetence or deliberate neglect of duty. Administrative problems can not be solved by evaluation. (See Chapter V.)

4. The evaluation problem chosen for study

The evaluation problem chosen for study will have to be one out of the complex of problems described under the preceding section, "The Problem Complex."

A good problem statement is one that is as concrete and specific as possible:

- In place of the total training effort of an institution or program, it may be preferable to evaluate a specific part of the training effort.
- In place of all aspects of a training effort, it may be preferable to evaluate only some aspects of a training effort.
- It may be preferable to cover a sample of a population rather than the total universe.
- It may be preferable to study the implementation of a training program during a specified time period rather than over the total life of the program.
- It may be preferable to look for specific and concrete effects of a training effort rather than its broad and generalized impact.

We are not suggesting that it is impossible or undesirable to study the broad impact of large scale training programs in terms of their general and long-term influences on large groups of trainees. All we are suggesting is that, in most situations, it might be more useful to be specific rather than general.

Whether the evaluation problem is defined in general or specific terms, ambiguity is not permissible under any circumstances. The evaluator, in stating his or her evaluation problem, should be most careful with words. The words should mean exactly what is in the mind of the evaluator, nothing more and nothing less, leaving no scope for alternative interpretations. (See Chapter IX below.)

5. Justifying the choice of the evaluation problem

The choice of one evaluation problem from a total "complex of problems" can not be capricious and arbitrary. The evaluator should be able to justify his or her choice of the particular evaluation problem.

The justifications may range from the political, to the programmatic, to the merely possible. An evaluation problem may be justified because the donors want it studied or because the planning department or the president's office needs to be pacified. At other times, the evaluation problem chosen may have important policy implications or may produce crucial feedback absolutely necessary for the future planning of a program. Or an evaluation problem may be justified in terms of feasibility -- something that can be accomplished with the minimum of resources even though there might be other more important evaluation questions which should have been tackled first if resources were available. (See Chapter V.)

6. Review of available research and experience

Available theory and research may help an evaluator to define and to clarify the evaluation problem and help in asking the right questions or frame the right hypotheses. Other evaluators, in other training settings, may have asked similar questions. Some experience may be available among administrators and trainers who have worked long in similar training situations. An attempt should be made to collect available knowledge, experience and opinion as part of developing the evaluation proposal. We should learn from other people's experience and should not waste our lives in reinventing the wheel!

7. Asking questions and sub-questions

It is important to translate the evaluation problem into a set of questions to be answered or hypotheses to be tested. Such questions may involve mere counting (How many of the literacy teachers trained during the last 5 years are still teaching literacy classes?); or describing (What is the nature of the emergent role of the family health field educator in Nakuru?). Hypotheses will involve relating (Training in cooperatives management is positively related to effective performance in the field); or predicting (Elementary school attendance in communities will go up as rates of adult literacy go up in those communities). Questions can be stated as hypotheses and vice versa. One need not, however, state one's evaluation interests both as questions and hypotheses, at the same one time. That will be a useless redundancy. It might be best to work with questions and sub-questions and leave hypotheses alone.

8. Evaluation model(s) and approach(es) to be used

Evaluators should give careful attention to the choice of evaluation models and approaches. At one level, the evaluator can choose between the classical (also called the scientific) paradigm or the naturalistic paradigm of evaluation and research. The choice between one or the other paradigm will lead later to different kinds of evaluation models and information-gathering approaches and techniques. (See Chapter VIII below.)

9. Evaluation design or steps and procedures

To have an evaluation design means to do all that is necessary to defend the conclusions of your study from attacks on validity and reliability. Quite often, the design may involve nothing more than the choice of right samples. At other times, evaluators may have to have base-line data or to have matched or control groups of some kind. Major steps in the conduct of the evaluation study and the procedures to be followed at each step should be outlined in this section of this proposal.

In the case of naturalistic evaluations, the problems of design take a different form and will be discussed in Chapter XII.

10. Instruments and tools of data collection

The proposal for an evaluation study should include a discussion of the tools and instruments that will be used for the collection of data. Preferably the first drafts of the tools and instruments should be attached to the proposal.

There are two prior questions that the evaluator must face before getting on with the construction of the tools and instruments: (1) What is the unit of analysis? In other words, where are effects and consequences likely to appear: in individuals, in families or groups, in organizations, or communities? (2) What will be the indicators of effects and consequences having actually appeared? In other words, what responses and behaviors, for example, will indicate change in motivations or in the learning of self-reliance? (See Chapter IX.)

The most frequently used tools and instruments in educational evaluation are:

- Tests
- Interviews
- Questionnaires, and
- Observation schedules.

Tests can perform a variety of evaluative functions. Achievement tests are like examinations to which we must have all been exposed during our years of schooling. Achievement tests measure achievement of trainees on the subject matter taught to them.

Aptitude tests measure the natural or acquired bent of mind. They are futuristic. They tell you whether a trainee is likely to make a good field worker or is he or she most likely to prefer a desk job? These tests are much more difficult to make than achievement tests.

Achievement tests may be pencil and paper tests or they may seek to measure actual performance on a task. When performance is the focus of measurement, these may be called performance tests. Performance tests are much more difficult to score than paper and pencil tests.

Finally, evaluators should understand the distinction between standardized tests and criterion-referenced tests. Standardized tests are typically used within formal instructional settings. Evaluators in developmental settings will most often be dealing with criterion-referenced tests: tests which determine and evaluate knowledge or performance in reference to criteria of success established by the trainer in relation to the particular trainee group, in the particular training setting.

Interviews may be structured (where you ask pre-determined questions, using exactly the words in which your questions have been written); or they may be unstructured (wherein questions are used merely to start a dialog; and further questions are framed and raised by the interviewer as he or she goes along probing and encouraging). Unstructured interviews are the tools of naturalistic evaluation.

We can have individual interviews or group interviews. We can use a form to record interviewee responses or a note book to record responses; we can use a tape recorder; or we may want to write a note on the interview only after the interview. (Some evaluators do this so as not to inhibit the natural flow of interactions between the interviewer and the interviewee).

Structured interviews, typically, use pre-designed forms. Each question is followed by some space in which the interviewer can record the answer. If the interviewee (rather than the interviewer) records the responses, we can call this form, a questionnaire. Questionnaires can be filled in person-to-person situations to have control on data collection. The evaluator can ensure that subjects sit down to fill in the questionnaires and the evaluator can take the responses back to the office. Questionnaires, however, can be very expensive to administer in face-to-face settings, if the respondents are spread all over a district, a province or a country. In that case such questionnaires must be mailed. Responses from mailed questionnaires seldom, if ever, show hundred per cent returns.

In some evaluation situations, important data may have to be collected through observation -- by being there and watching and listening. Observation, again, as in the case of interviews may be structured or unstructured. Structured observations will, of course, require pre-designed observation schedules. Unstructured observations may be recorded in a note book or spoken into a tape recorder by the observer. Observations may be conducted by a single evaluator or by a team of evaluators. Again, observations may be done at a single time, or serially, by returning to the same situation at different intervals of time. We can even apply sampling techniques to observation by making samples, random or otherwise, of places to be visited, of days of visitation and of time units when observation will be conducted.

Observations may sometimes be done obstrusively (subjects of evaluation will know that they are being observed) or unobstrusively (subjects of evaluation will be unaware of being observed). This later type of observation is the special tool of naturalistic evaluation. Even while doing unobstrusive observations, one could be a detached observer or a participant observer.

Evaluators have to be careful about the choice of tools and instruments in their data collection. Each tool has its strengths and weaknesses and each delivers a somewhat different kind and quality of data.

Finally, a suggestion that should be followed whenever possible: pretest the tools and instruments and data collection approaches in pilot settings. Rehearsals are as important for the act of data collection as they are for the acts of a play. (See Chapter X.)

11. Field work and related research plans

Proposal for an evaluation study should include plans for library research as well as data collection from the field.

If some documents or reports will be needed, the evaluator should know where to find them, who will have them, how to obtain copies of those documents and how much time, it might require for obtaining those materials.

Plans for collection of field data should be made carefully. If the evaluator can not collect all the data personally, investigators or interviewers may have to be hired. This means that plans must be made about their recruitment and training. Local contacts in the field must be identified and orientation must be provided to them about the objectives of research and about research plans.

Field visits must fit the realities of the field and the convenience of individual respondents. The evaluator must keep in mind such considerations as the harvesting season, the weather, fairs and festivals and visits of V.I.P.'s, examination schedules in schools and training institutions, and planning and budgeting cycles in departments and ministries.

Problems of transportation should be all anticipated and solved. Keeping all of the preceding in view, a time schedule should be prepared. (See Chapter XIII.)

12. Plans for data analysis

Data analysis plans must also form part of the proposal for the evaluation study. Will coding sheets or tabulations be needed for data collection? If so, these should be prepared and tested. Personnel needed for coding and collating data should be recruited and trained. Need for technical consultancy or statistical help (even computor time, if required) should be anticipated and plans made for receiving such help.

As in the case of data collection plans, data analysis plans must also be prepared in terms of a time schedule. Mere lists of things to be done is not enough; plans must be time-sensitive. (See Chapters XIV and XV.)

13. Budgetary plans

Conduct of an evaluation study will need staff time; secretarial and duplication help; paper, postage, tape and tape recorders (in some cases); field investigators; and transportation and telephone costs, etc.

All these resources exist within training institutions and programs and are available to those who want to use them. It is impossible to think of a training institution that would not want its trainers to do the best training job possible. Good training requires feedback; and, therefore, evaluation is an integral part of all good training. The resources available in the institution for "training" should be equally available for the "evaluation of training."

It can be said that the problem of resources is often self-created. Trainers may look at evaluation of training as something separate from training and, therefore, as an extra burden. When asked to do an evaluation of training they may ask for extra time, extra resources and extra credit for the evaluation work.

The new orientation towards evaluation as part of training, will not, however, solve all resource problems. Institutions may not have budgeted for transport, postage, stationary and secretarial help to include the institution's obligations to evaluate its own programs. In such cases, extra budgetary provisions will have to be made and resources needed for conducting the evaluation study will have to be obtained.

14. Report writing

The proposal for an evaluation study should also include the element of "reporting plans." Will the evaluation results be used within the program or the institution or will those be disseminated outside the institution? If dissemination outside the institution is envisaged, a clear description of outside clients and consumers of the evaluation study should be developed. The same report is not necessarily appropriate for all groups; and writing different versions of the report should be considered.

In writing an evaluation report, the policy and program implications of data should be brought out. Data does not always speak for itself. While it is necessary that evaluators bring out the implications of their findings for policy makers and program planners, they should not draw unwarranted conclusions. Opinions and hunches should not be mixed with inferences from the data.

Evaluative information can be both used and abused. Too often readers of evaluation studies may be in search of culprits rather than causes; and may want to punish rather than plan with greater understanding in the future. No wonder that colleagues whose work is being evaluated will often get worried about the evaluation process and what it might find. To handle the departmental politics of evaluation, it may be useful to discuss the preliminary report of evaluation in a group setting before issuing a final evaluation report.

All evaluation studies need not be duplicated and distributed. A single copy of an evaluation study will be worth a thousand, if its findings illuminate action and if its recommendations become part of decision making. (See Chapter XVI.)

15. Bibliography

A proposal for an evaluation study should also include a bibliography of books, reports and documents used in developing the proposal and likely to be used in the conduct of the study and in writing the final report.

Summary

The time and effort given to writing formal proposals for evaluation studies is often well spent. Elements in an evaluation proposal may include (1) the development context of the training program; (2) the description of the training program in design terms; (3) the problem complex; (4) the evaluation problem chosen for study; (5) justifications for choice; (6) review of available knowledge on the evaluation topic; (7) questions and sub-questions to be answered; (8) the evaluation approach selected; (9) design, steps and procedures; (10) instruments and tools; (11) plans for data collection; (12) plans for data analysis; (13) budgetary plans; (14) reporting plans; and (15) bibliography and list of resources.

Things to Do or Think About

1. Prepare a formal proposal for an evaluation study of your choice.
2. Have you conducted an evaluation study before? Do you think your evaluation study could have been improved if a formal proposal had been

written before the actual implementation of the study? If you have never conducted an evaluation study yourself, discuss the usefulness of the ideas included in this Chapter with someone who has.

CHAPTER VII

MONITORING AND QUICK APPRAISALS

There has been considerable interest recently in development literature on strategies for gathering quick evaluative feedback on the performance of programs. The point is made that typical evaluation studies may too often take too long a time for decision makers to wait for results. Program decisions will often demand quick pulse-taking of programs to get a report card on the health of programs. Time is often the essence.

At least three different strategies can be suggested for such quick pulse-taking:

(1) Monitoring,

(2) Quick Appraisals, and

(3) Networks of Informants.

The question we must ask here is: Do monitoring, appraising, and network building also require written proposals much like those suggested in the previous chapter? The answer can not be given in a simple 'Yes' or 'No'. For example, a different sort of preparation will be needed for establishing a monitoring system which almost always will be built upon an adequate management information system (MIS). The development of an MIS, in turn, will require extensive planning and development of proformas for recording and reporting program implementation data. Quick appraisals, as we will see later, may require less exhaustive written proposals than the ones suggested in Chapter VI, but will yet demand considerable formal preparation. Planning for networks of informants will also require forethought and preparation.

(1) Monitoring

To monitor is to watch or check upon an on-going program to detect flaws and to provide corrective actions.

Monitoring (or Concurrent Monitoring) is typically based on data being routinely generated by a program in the very process of implementation at its various locations, levels and phases.

Monitoring is done best when assisted by a management information system (MIS). An MIS makes a program more monitorable and, thereby, more manageable.

The following considerations must enter the design of an MIS:

(i) Choice of units and entities where change will appear. Such units may be individuals (learners, trainers, community leaders), groups (women's clubs, discussion groups), institutions (health clinics, rehabilitation centers), or communities or sub-cultures. Sometimes these may be physical entities, such as, homes, fields, shops, wells, storage bins, etc.

(ii) Choice of indicators of change. Indicators or signs that signify change in already selected units and entities must be chosen next. These indicators may be attendance in literacy classes, absence from the factory for reasons of ill-health, purchase of consumer goods, rise in productivity, etc.

(iii) Economy in data collection. In designing an MIS one need not collect all possible data. Indeed only the minimum necessary information should become part of the MIS.

(iv) Selection of points for origination, successive aggregation and storage of data. It should be clear where data will originate, at what successive levels and points data will be aggregated and where it will be finally stored for possible retrieval and use in decision-making. This will require that data collection and data aggregation duties of various role performers in the program system are clearly identified; and the logistics of the physical transfer of data are managed effectively.

(v) Design of proformas, tabulations, report forms, etc. Simplicity and standardization of instruments for data collection and reporting are important for the successful implementation of an MIS.

(vi) Timing and flow of data. Monitoring fails if it is not time-sensitive. Data should move in time series according to a determined pattern of flow -- upwards, downwards and horizontally as required.

(vii) Data audit. Data audits are important to ensure that there is no misreporting of data and that the intent of various proformas in use in the MIS is well-understood.

(viii) Processing and feedback. The data collected at a particular level should be used to improve program decisions at that level before sending it upwards. When data is processed at the central level, or at one of the intermediate levels, feedback should be provided to all concerned. Only then can the total system grow in intelligence.

(2) Quick Appraisals

Quick appraisals are quick evaluations, conducted under conditions of emergency to investigate the cause of a breakdown, to anticipate problems, or to get early returns on the impact of a program.

Quick appraisals may be less exhaustive and less comprehensive than regular evaluation studies, but one needs to prepare for a quick appraisal carefully and systematically. A short form of the proposal as discussed in Chapter VI should be prepared.

Appraisal teams will have to be carefully built, and given a clear mandate regarding the information they should collect and the judgments they should render. The team should think about the mix of the quantitative and qualitative data they will try to collect.

The instruments to be used for the quick appraisal should be carefully designed and pre-tested. Samples of respondents should be small but carefully chosen.

Dead-lines are important to keep in the case of quick appraisals, otherwise, they will not be quick appraisals any more. A time-frame of 4 to 6 weeks is typical for quick appraisals.

A quick appraisal is more than monitoring which uses routinized data generated in the very process of the implementation of a program. Quick appraisals, on the other hand, do collect fresh data for the purpose of answering significant questions. However, whenever possible, quick appraisals should make secondary use of data already stored in the management information system.

Self-reports by functionaries of a program on their own performance is another important resource for quick appraisal teams. Similarly, self-reports by the beneficiaries of a program should be used for developing quick appraisals.

(3) Adapting Mass Observation to Networks of Informants

Professor Madge of Great Britain developed an interesting feedback technique that he has called "Mass Observation." The mass observation technique uses untrained volunteer reporters who make their observations in the normal course of their lives and in the natural course of programs and events. Then they describe in writing, fully, clearly and simply what they have observed and what they think their observation add up to. Such reports may include photographs, clippings from the local newspaper, things heard over the local radio station, or overheard in the local bar. Sometimes trained interviewers may also be added to the corps of people observers. Advocates of this technique believe that it is easier to detect the bias of an untrained observer than that of a social scientist and suggest that mass observation should not be dismissed as a biased approach to the collection of feedback. They assert that it is an excellent approach to the study of the "mind life" of the people.

The mass observation approach can be adapted to developing special networks of informers who do a limited version of the mass observation for the evaluators. A whole group of people located at various places and levels of a program can be asked to observe and write about a program, its process, its impact and consequences.

Summary

Urgency of the need for feedback may demand quick pulse-taking rather than systematic evaluation of the program. Three different approaches can be used for generating such feedback: (1) monitoring, (2) quick appraisals, and (3) networks of informants. Each of these approaches will require adequate preparation, even though formal proposal writing as suggested in Chapter VI may not be undertaken.

Things to Do or Think About

1. Can your training center or institute make use of an MIS? What might be some of the essential elements of such an MIS?

2. List the names, official designations or occupations of at least ten people you would include in your network of informants to collect feedback on your program. What special contribution you expect each of them to make?

CHAPTER VIII

MODELS OF EVALUATION

In the literature on evaluation, one is likely to come across the terms 'evaluation approaches', 'evaluation models' and evaluation paradigms'. An understanding of the meanings of these terms is essential for a thorough comprehension of the ideas presented in this chapter.

Let us begin with the word 'paradigm'. In the dictionary meanings of the word, a paradigm is an ordered list, a table of classes, a pattern, or a formula for the general form into which specifics of a certain order may be placed.

Kuhn[1] in his study of scientific revolutions gave a special meaning to the word paradigm. He defined a paradigm as the "creative ideology" of scientists from which they worked, and which provided them with a particular logical and methodological stance for producing scientific or social scientific knowledge.

Thus, evaluation paradigms are the creative ideologies of evaluators. These paradigms determine the thinking and methodological behaviors of evaluators: what they think about the nature of reality; and how they think dependable statements can be made about stability and change in the social reality that surrounds us. There are two basic paradigms of evaluation (and research) that we will be discussing later: the scientific paradigm and the naturalistic paradigm.

---

[1] T. S. Kuhn, The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1962.

We should define models next. A model is information, data or principles grouped, verbally or graphically (and sometimes mathematically) to represent or describe a certain thing, idea, condition or phenomena. A model is the essence of the learning and thinking of a specialist, stated clearly and parsimoniously. Models are the progeny of paradigms. It is important to remember this relationship between paradigms and models.

Evaluation models are thus verbal, tabular or graphic presentations of the principles learned by evaluators. They are the essence of their separate experiences developed in the background of particular paradigms.

Finally, an approach is merely a method of beginning or accomplishing something. However, when an approach, through successive use and testing, becomes both standardized and formalized, it acquires the status of a model.[1]

Two Basic Paradigms of Evaluation in Education and Development

There are two basic paradigms of evaluation in education and development:

1. The Scientific Paradigm of Evaluation, and
2. The Naturalistic Paradigm of Evaluation.

---

[1] In our discussion in this chapter, the term 'theory' has not been defined or explained. This is so because the literature on evaluation talks often of evaluation models and seldom of evaluation theory. Let us say briefly that, in terms of a conceptual status, theory falls between the paradigm and the model. In its best sense, theory is a deductively connected set of laws and empirical generalizations. A model is often a schematic diagram that connects theory with practice.

The scientific paradigm

The scientific paradigm is so called because it copies the methods of hard sciences: physics, chemistry, botany, engineering, etc.

It is sometimes called the classical paradigm because it is old, follows strict rules and is seen as standard and authoritative.

It has also been referred to as the positivist paradigm because of its emphasis on the explicit; that which is capable of being directly and certainly affirmed.

The ideal of the scientific paradigm is the randomized controlled experiment. (Quasi-experimental designs may be acceptable under some conditions.) The scientific paradigm demands a clear definition of evaluation objectives, of variables, a sampling plan, structured instrumentation that generates quantitative data, statistical techniques in the analysis of data and generalizability of results.

The naturalistic paradigm

The naturalistic paradigm suggests that human behavior be studied in natural settings, within its total context. It is the method of the anthropologist and the ethnographer. It is holistic - it studies reality whole, without dividing it artificially into parts and segments.

The naturalistic paradigm is also referred to as qualitative and phenomenological. This means that, unlike the positivist, the naturalistic evaluator seeks to describe phenomena and search for regularities and patterns. He searches for understandings of the specific situation that may later illuminate other somewhat similar situations. He does not search for generalizable laws.

In naturalistic inquiry, the design is emergent; it emerges as the evaluator undertakes different steps and procedures in the collection of meaningful data. The samples are purposeful rather than random. The instruments are always unstructured and generate qualitative data. Claims are made in regard to the applicability and fittingness of results rather than to their generalizability.

The scientific paradigm versus the naturalistic paradigm

The scientific paradigm has had great victories in the hard sciences. It has produced research that has banished diseases from the face of the earth and has put man on the moon.

It was so successful that social scientists (sociologists, psychologists, economists, educators, even anthropologists), also wanted to use the scientific paradigm of the physicist and botanist. They used it with a vengence. It made them feel like real scientists. For years and years, the scientific method was learned and the scientific method was taught in all the social science departments of universities.

The realization has emerged during the last twenty years or so that the scientific paradigm has given social scientists good feelings but not good scientific results. We have discovered that too often social life does not fit into the laboratory. In trying to control variables, we segment human behavior unnaturally and indeed change the very nature of the human behavior being studied. Aggregation of scores and statistical treatments of data may look elegent and impressive but results may be trivial and even misleading. In human behavior the context is important.

We need to study not just behavior but behavior-in-context.

The naturalistic paradigm is more appropriate, most of the time, for the study of human behavior. Once rejected out of hand as subjective and qualitative, it is becoming more and more acceptable. As its methodology becomes clearer and techniques of data analysis are further advanced, the naturalistic paradigm of evaluation will find its rightful place in evaluation methodology.

The readers of this book should not select one of the two paradigms discussed above and reject the other. They should learn to consider them to be complementary. The scientific paradigm is necessary when developing profiles of large-scale programs for use by decision makers at the center. The naturalistic paradigm is needed when the emphasis is on an understanding of processes and effects at program levels. Typically, educational evaluators will be using both these paradigms within the context of the same one evaluation study.

The following table summarizes the differences between the two paradigms of evaluation:

TABLE 1. DIFFERENCES BETWEEN THE TWO BASIC PARADIGMS OF EVALUATION

The Scientific Paradigm	The Naturalistic Paradigm
Philosophic roots	
Positivist, reductionist	phenomenological, holistic
Design	
Experimental or quasi-experimental	Emergent design (or rolling design)

Setting of evaluation/research	
Laboratory or otherwise controlled	Ecological; in context
Sampling	
Random. Size pre-determined	Purpositive, elite, specialized. Size determined in use; sample is exhausted when available information is exhausted
Methodological orientation	
Quantitative	Qualitative, thick description
Instrumentation	
Structured, often interventionist. Instruments are sought to be standardized and made independent of evaluator	Unstructured, often unobstrusive. Researcher/evaluator himself or herself becomes the tool of data collection
Data analysis	
Statistical	Thematic; content analysis of interviews, documents, and observations
Report	
Statistical-analytical	Descriptive, interpretive; typically a case study
Strengths	
Provides good estimates of differences, variations, and correlations when variables can indeed be properly defined and reasonable controls can be established	Responsive, adaptability, holistic emphasis

Guba and Lincoln in their recent book[1] have discussed the various aspects of rigor as they apply to the two basic paradigms of evaluation. What they are saying is that both the scientific and the naturalistic paradigms of evaluation can be rigorous in their own ways. See table below.

---

[1] Egon G. Guba and Yvonna S. Lincoln, Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches. San Francisco, Ca.: Jossey-Bass, 1981. (page 104)

TABLE 2. SCIENTIFIC AND NATURALISTIC TERMS APPROPRIATE TO VARIOUS ASPECTS OF RIGOR

Aspect	Scientific Term	Naturalistic Term
Truth value	Internal validity	Credibility
Applicability	External validity/ generalizability	Fittingness
Consistency	Reliability	Auditability
Neutrality	Objectivity	Confirmability

Models of Evaluation

Against the background of these two general paradigms, many different models of evaluation have been proposed by specialists in the field.

Why are there so many evaluation models? Isn't there one correct way of doing evaluation?

Earlier in this chapter, we defined a model as the essence of the learning and thinking of a specialist, stated clearly and parsimoniously for communication among professionals and practitioners. There are many different evaluation models, because different specialists have undergone somewhat different learning and thinking. They have accepted one basic evaluation paradigm or the other. More and more evaluation specialists now suggest a pragmatic mix of the two paradigms - the scientific and the naturalistic.

Evaluation models are different also because they have emerged within different program settings: within formal education or within out-of-school and nonformal education settings; within mental health setting in an industrialized country or within family life education in the context of a developing country.

Finally, and most importantly, evaluation models are different because evaluation specialist have introduced additional "value" considerations to their initial choices of paradigms. Some evaluation models emphasize a more synoptic view of evaluation, suggesting that we evaluate not only the behavior of our so-called clients but also of our own. Some evaluation models suggest the introduction of imagination to our evaluations and not depend only on cold calculation. Some suggest that the unanticipated consequences of program actions may be as important as the intended and the anticipated. Therefore, the model of evaluation should be able to accommodate both the anticipated and the unanticipated consequence. Some suggest evaluation to be conducted as an advocacy and confrontation. Some suggest participative evaluation wherein both the means and ends of evaluation are participatively determined by all concerned - organizers, professionals, and beneficiaries.

One can see a clear underlying value direction in the development of evaluation models during the last twenty years: (1) there is exclusive or complementary use of naturalistic strategies; and (2) there is a move towards inclusion of the beneficiaries of programs in the design and implementation of evaluations. The key words are holistic and participative.

Some of the evaluation models often referred to in the literature of evaluation will be discussed below. The discussions will be brief. We include a discussion of the evaluation models for two reasons: educational and political. The trainer-evaluator should have some idea of what different evaluation models exist and what their characteristics are. This is for his or her education. But there is also a political reason. The trainer-evaluator should be able to justify his or her choices of the model or models; and should be able to hold his or her own against the outside specialist. We should not allow technicians and specialists to browbeat us with the use of unfamiliar names and phrases.

Before we proceed with short descriptions of evaluation models, we should caution readers against a search for the model, for one correct way of doing evaluation. As in the case of paradigms, a pragmatic mixture of models may be often the best thing to try.

The Context-Input-Process-Product (CIPP) Model of Evaluation

This model is often associated with the name of Daniel L. Stufflebeam who has used this model in various evaluation studies.

According to this model, the sole purpose of evaluation is to produce information for decision makers. Using the system metaphor and the four parameters of systems (context, input, process and output), the model talks of four types of evaluations to provide information for four types of decisions:

1. Context evaluation - to provide information on the setting, to be able to make planning decisions;
2. Input evaluation - to make programming decisions;
3. Process evaluation - to make decisions related to methodologies and implementation; and

4. Product evaluation - to evaluate impact and to make recycling decisions.

The CIPP model comes closest to formal research models with its stress on the clarification of evaluation objectives, structured observation, and the testing tradition of achievement testing in schools. The model adopts the criteria of internal and external validity, reliability, objectivity, relevance, importance, scope, credibility, timeliness, pervasiveness and efficiency of the evaluative information produced. It has, however, shown little concern for values. The model suggests that evaluators should produce reports that decision makers can use in making decisions.

The Countenance Model of Evaluation

The countenance model of evaluation is associated with the name of Robert E. Stake. It is so called because Stake talked of two countenances of evaluation - description and judgment.

This model was directly related to the evaluation of effects in terms of stated objectives and involves the completion of two data matrices as follows:

	Description Matrix		Judgment Matrix	
	Intents	Observations	Standards	Judgments
Antecedents				
Transactions				
Outcomes				

The task of the evaluator is to find data for all the cells in the table on the previous page: to compare observations to intents: and to make judgments in terms of the standards agreed to among program organizers and evaluators. One should note that in systems vocabulary antecedents are inputs and transactions are processes.

The model in implementation has used stratified random samples for collecting special information, combined with the case study approach.

The model has called attention of evaluators to the need for defining standards on the basis of which judgments can be made, though the model itself has left the question of specification of standards unresolved.

The Discrepancy Evaluation Model

The model was proposed by Malcolm Provus who defined evaluation as the art of describing a discrepancy between expectation and performance of a program.

The model suggests that we look for discrepancies in terms of five different aspects of a program:

1. the design of the program
2. its installation
3. the processes of implementation
4. the product, and
5. the cost

While recognizing the usefulness of the experimental method in certain cases, the model shows preference for the descriptive methods of history and anthropology and the case study method of sociology and psychiatry. Thus, the model accommodates both the scientific and the naturalistic paradigms.

With its relative emphasis on naturalistic methods, it suggests that evaluators work in teams to be able to test individual perceptions of each against the other and to be able to question the standards being applied to describe discrepancies.

The model claims to provide continuous information to decision makers on the performance of an on-going program. It also claims to provide information that has direct one-to-one relationship to decisions actually being made. The resources required for effective application of this model can be considerable in terms of personnel, time and money.

The Transactional Evaluation Model

The transactional evaluation model is associated with the name of Robert M. Rippey who has challenged educators and trainers to show what they have accomplished rather than what scores their students and trainees have made.

The focus is on educational accountability - change makers are asked to study themselves, their roles, the systems in which they play these roles and the larger systems that surround the systems under change.

In Rippey's[1] own words:

> A comparison with traditional summative and formative evaluations show that the target of evaluation is different: the subject of evaluation is the system, not the client or the services rendered by the system. The variables relate to the social, psychological and communication aspects of the system, rather than to the manifest objectives. The information is continuously fed back into the system. The evaluator himself is more a part of the operating system. The conventional considerations of reliability, validity and objectivity are less important than those of timeliness, relevance and the observable effects of generating evaluation information. Primarily, evaluation is intended to transform the conflict

---

[1] Robert M. Rippey, ed., Studies in Transactional Evaluation. Berkeley, Ca.: McCutchan, 1973, pages 3-4.

energy of change into productive activity; to clarify the roles of those persons involved in the program changes, not to produce new knowledge or ascribe causality.

One should note the assumptions in regard to the basic paradigm in use in the transactional evaluation model and the additional value positions introduced in the model. It is indeed a highly value-laden model. It emphasizes relational information and urges sensitivity to the unanticipated consequences. It also implies that evaluation be conducted collectively by the protagonists and designers of a change program and representatives of those likely to be affected.

The Goal-Free Evaluation Model

The idea of goal-free evaluation was introduced by Michael Scriven. He pointed out that in our emphasis on stated goals, our search got completely focussed on intended effects - effects we wanted to create under accepted program goals. This focus became so exclusive that we often developed a tunnel vision: looking for evidence of intended effects and seeing nothing else.

He suggested that we should look for the real effects of programs - effects that had actually occurred whether intended or unintended. This he thought could be done if we conceived of a goal-free evaluation, independent of objectives stated for programs. Results from objectives-focused evaluation and goal-free evaluation of a program could then be combined.

The Investigative Social Research Model

Jack D. Douglas[1] has recently analyzed the methods of the investigator and show how investigative strategies can be used to expose the truth about people in social settings

---

[1] Jack D. Douglas, Investigative Social Research. Beverly Hills, Ca.: Sage Publications, 1976.

The investigative model does not assume a world of cooperation, openness and truthfulness, but one of misinformation, evasions, lies and fronts. He then suggests strategies for grasping an evaluation setting, infiltrating the setting, building friendly and trusting relationships, and then using them in a continuous process of testing out and checking out.

The _modus operandi_ model, suggested by Michael Scriven is also an investigative method for studying cause-effect relationships through sequential testing. This method reconstructs "the procedures of the historian, the detective, the anthropologist, and the engineering troubleshooter." The _modus operandi_ model is proposed as a substitute for experimental and quasi-experimental approaches when field situations preclude their use. Essentially, the method involves generating hypothetical chains of cause-effect events and eliminating those that could not possibly have happened. This is the typical method of the detective.

"_Evaluation as Illumination_"

This model was developed in clear rejection of the "agricultural-botany" model of evaluation rooted in the scientific paradigm. It was asserted that groups and communities can not be randomly assigned to treatments as farms and fields; and human beings can not be administered treatments as seeds in the ground. In any case, quantitative data generated by the agricultural-botany model provided only partial descriptions of phenomena.

Parlett and Hamilton[1] built their model on two important considerations.

1. Instructional systems, once adopted, become living systems. Living systems do not match their catalog descriptions. Important modifications occur in programs as they move from the drawing board to actual implementation.

2. Programs of training and development can not be separated from their learning milieu. Actors in the learning milieu and the structures of the milieu become part of the instructional system.

While retaining the use of sampling methods, and structured questionnaires and tests, Parlett and Hamilton drew our attention to the naturalistic methods for description and interpretation. Three stages in the evaluation process are suggested to include: (a) observation of the educational setting; (b) selection of themes through progressive focusing and intensive inquiry; and (c) analysis and explanation.

The Connoisseurship Model of Evaluation

The connoisseurship model of evaluation proposed by Elliot W. Eisner[2] makes a clean break with the scientific paradigm and draws from the aesthetic tradition of the arts.

Eisner suggests two interrelated concepts: (a) educational connoisseurship and (2) educational criticism to perform the tasks of educational evaluation. Educational connoisseurship is the "means through which the shape of the context and the configurations within it can be reorganized so that intelligent decisions about the context can be made." Educational criticism is the art of disclosure through description, interpretation and evaluation.

---

[1] M. Parlett and D. Hamilton, "Evaluation as Illumination: A New Approach to the Study of Innovatory Programs." Occasional Paper, No. 9. Edinburgh: Center for Research in the Educational Sciences, University of Edinburgh, 1972.

[2] Elliot W. Eisner, *Educational Imagination: The Design and Evaluation of School Programs*. New York: Macmillan Publishing Co., 1979.

The methodology of connoisseurship and criticism is by no means soft-headed or romantic, and certainly can be systematic and rigorous. Educational critics can learn to look for the pervasive qualities of education in the classrooms and training settings; and can learn to look for the meanings of hidden cues.

Questions of reliability and validity must be handled through structural corroboration (mutual validation of bits of data one by the rest; and the whole being supported by the bits that constitute it); and referential adequacy (the existence of relationship between what the educational critic says and the subject matter of his or her critique). Generalizations are also possible in the sense that educational criticism will lead to more refined processes of perception in subsequent settings; and will create in the evaluator's mind new anticipations.

Reports of educational criticisms have a family resemblance to case studies, but case studies of educational criticism are different in the sense that criticism itself is an art form. As a critical disclosure, educational criticism report creates a living image, communicating to its readers a visceral understanding of the educational realities.

The Advocacy Model of Evaluation

The advocacy model is given this name since it allows groups of people both for and against a program to advocate their position before an educational jury. This model is, thus, evaluation in the legal mode. It is an educational trial by jury.

An educational jury is established and groups with opposing views on the usefulness of a program argue their case before the jury. All kinds of data quantitative and qualitative can be presented and any witnesses may be called to testify. These witnesses may include specialists and non-specialists, organizers and beneficiaries, men and women, and adults and children.

The model seeks to bring out both sides of an issue in a balanced fashion, but most of its advantages are counterbalanced by corresponding disadvantages. The model introduces controversy and confrontation where none may have existed. Good juries are hard to put together. The format does, however, force administrators to consider that there may be some things wrong with the on-going-program and to consider the possibility of discontinuation or drastic transformation of a program. An important advantage of the model is that evaluative information is used to come to concrete decisions in regard to the continuation or re-design of programs.

The Participative Evaluation Model

The name of Paulo Freire, the Brazilian educator and the author of Pedagogy of the Opressed (Herder and Herder, New York, 1972) is often associated with participatory evaluation and research. Considerable amount of work has since been done in this area during the last ten years by evaluators spread all over the world. Participatory research networks

---

[1] Participatory Research: Developments and Issues. A special issue of Convergence, Vol. XIV, No, 3, 1981.

have been established, participatory research and evaluation studies have been conducted and their results published.

Participatory research or evaluation is not a scientific endeavor of the professionals, but an in-depth, existential review of an experience done by all concerned, together in mutual collaboration. The learner becomes an evaluator and the evaluator becomes a learner. Evaluation goals, ends, standards and tools are decided upon participatively. Each contributes personal data and collects the data that has to be obtained. Analysis of data is collectively undertaken. Judgments are also rendered collectively.

In an address to the Institute of Adult Education, University of Dar-es-Salaam, on 20 July 1972, Paulo Freire presented the possible steps in such a participative methodology:

1. The evaluation (or research) team should acquaint itself with all previous research and evaluation - no matter what methods were used in that research evaluation.
2. The team should delimit the area of action geographically - even though, culturally speaking, there are no frontiers.
3. The team should identify official and popular institutions in the area selected and go to talk to the leaders within those institutions.
4. The evaluation team should tell these leaders, in all honesty, that they have come to discuss the possibility of all people in that community holding discussions and working together.
5. If the leaders agree, the evaluation team should hold meetings not only with the leaders of various institutions but also with the people who are involved in some way with those institutions.

6. The evaluation team should discuss with the community arrangements for meetings wherein groups of, say, thirty people could come together on a daily or weekly basis for discussions. Such meetings might involve almost all the inhabitants of a community and last for several weeks. The important thing would be to obtain a perception of the whole community.

7. Sociologists, psychologists, educators and linguists should, at this stage, join the research or evaluation team and visit each group. Records of discussions should be made at each meeting. People should be urged to speak if they are silent, but otherwise the role of the evaluation team should be no more than advisory. One of the members of the community should chair such meetings.

8. Justice, education, government, industry and many other topics may be discussed; but all in terms of the people and in the context of concrete realities.

9. When the smaller groups think they have exhausted the topics for discussion, each one should put its findings on paper and then they should all meet in a general session. The reporters at such sessions should be the people themselves; not the specialists on the team. The workers should become intellectuals. There should be collective discussion of each group report.

10. The evaluation team should now make a critical study of the people's discourse. This study should be interdisciplinary. The various levels at which people perceive reality must be determined and their many implications should be worked out. These implications must be studied in the presence of the people, not by social scientists on their own.

11. The evaluation team together with the people should now draft a proposal for subsequent action. The programme itself should not be worked out for the people but with the people.

It should be clear from the preceding that participative evaluation is not distinguishable from need assessment or community awareness. The distinction between evaluation and instruction gets lost in participative evaluation. Participative evaluation provides participants with further opportunities to raise their consciousness and consolidate their sense of power and self-worth.

The Situation-Specific Strategy (3-S) Model of Evaluation

Before presenting the 3-S evaluation model, let us remind our readers that it is useless to look for the model of evaluation, or for one correct way of evaluating training or development. A trainer-evaluator might often be using more than one of the above models, within the context of a single evaluation study.

Another important point to remember is that models are not usable as formulas. All of your problems and needs will seldom fit neatly into an evaluation model . Models should indeed be used to "think with", not as step by step and unchangeable sets of procedures.

The 3-S model to be discussed below is a model that should help us select a good mix of models and approaches to be used in an evaluation study. The conceptual essence of the 3-S model is this: Do not start with the evaluation model, begin with the evaluation problem. Analyze the evaluation problem into sub-problems; think how the problem or parts of the problem might unfold over time; and, finally, think of the milieu in which evaluation will be conducted.

Different parts of the evaluation problem will most likely require different evaluation models and approaches. You may need both a survey and an in-depth case study. You may require achievement testing as well as content analysis of documents.

The exigencies of time may demand pulse-taking through quick appraisals, even though, ideally, a more systematic evaluation would have been better. Finally, the evaluator may be working in a situation where there are no calculators or colleagues who can help with the analysis of large bodies of numerical data; where there are no copying machines or stencil duplicators; or where there is no duplicating paper for producing the required instruments. The 3-S model helps us think about what strategies to choose in specific real-life situations. How to do the second best when the best is not possible?

Elsewhere,[1] we have listed the following steps in the implementation of the 3-S evaluation model:

1. Ordering the world of change and evaluation (Chapters II and III)
2. Articulating the means-ends relationships in the change program (Chapter IV)
3. Generating profiles of information needs and evaluation issues (Chapter V)
4. Developing a situation-specific evaluation agenda (Chapter V), and
5. Choosing appropriate and realistic methodologies and techniques (Chapter VIII and subsequent chapters).

---

[1] H. S. Bhola, Evaluating Functional Literacy. Amersham, Bucks, U.K.: Hulton Educational Publications Ltd., 1979. Pages 25-33.

Summary

There are two basic paradigms of research and evaluation: (1) the scientific paradigm and (2) the naturalistic paradigm. Against, the background of these two basic paradigms, a variety of evaluation models have been proposed, among them: the CIPP model, the countenance model, the discrepancy model, the transactional model, the goal-free model, the investigative model and the *modus operandi* approach, evaluation as illumination, the connoisseurship model, the advocacy model, the participative model and the situation-specific strategy (3-S) model. The trainer-evaluator has to understand the strengths and weaknesses of the various paradigms, models and approaches and learn to develop conceptual strategies that fit his or her problems in the particular context.

Things to Do or Think About

1. Of the two basic paradigms discussed in this chapter, which is likely to generate more useful information on your training program? Or, do you have to use both?
2. Of the models described in this chapter, which model or models do you personally consider most useful in your work? What more would you like to know about the model to put it to work?
3. Can you find evaluation studies already completed that fit neatly under one or the other model described in this chapter?

CHAPTER IX

CONCEPT ANALYSIS, INDICATORS OF CHANGE AND STANDARDS OF QUALITY

In an earlier chapter, we talked of raising evaluation questions, of developing evaluation agendas and about typical evaluation themes of interest to trainers-evaluators. Let us examine here, some of the evaluation questions that trainers-evaluators have actually asked themselves and the evaluation topics that they have set out to study:[1]

1. To determine the relative effectiveness of different sets of post-literacy teaching-learning materials being used in the programs for new literates, in terms of content, occupational orientation, style and format.
2. An evaluation of the efficiency and effectiveness of the "Administrative Skills Courses for Chiefs" offered by the Government Training Institute (GTI), Maseno, Kenya, during 1973-77.
3. An evaluation of the basic education curricula for the Undugu schools in Nairobi.
4. An evaluation of the resettlement programs for the vocationally-trained disabled persons in terms of their social and economic rehabilitation.
5. An evaluation of the effectiveness of the Form IV Literature-in-English radio program in the Central Province of Kenya.
6. An evaluation of the suitability and the utilization of available health education materials by the Family Health Educators in Nakuru; and an exploration of the need for producing materials suitable for local use.

---

[1] These examples have been taken from the evaluation questions and topics selected by the participants of May 1979 and August 1980 workshops on the Evaluation of Basic Education and Development Training Programs, Kericho, Kenya.

7. Reasons for dropouts from the diploma education training program at the National Institute of Education, Makerere, Uganda.

8. To evaluate the patterns and frequency of library use by students at Machakos Teachers' College, Kenya.

9. Evaluation of the effectiveness of the Nutrition Surveillance Curriculum for the July 1980-March 1981 class in the Nutrition Course at Karen College, Kenya.

10. How effective is the utilization of handicraft skills given to Field Workers in improving the nutrition status of the peri-urban communities that these field workers serve in the Nairobi area.

The preceding examples provide us with a small sample of the types of evaluation questions that can be asked by trainers-evaluators. But in this chapter, we are interested in the words and phrases used in these various statements of evaluation topics. Let us examine the words and phrases which have been underlined in the examples above:

determine (example, 1)
relative (example, 1)
effectiveness (examples, 1, 2, 5, 9)
new literates (example, 1)
content (example, 1)
format (example, 1)
efficiency (example, 2)
basic education (example, 3)
curricula (example, 3)
disabled person (example, 4)
social and economic rehabilitation (example, 4)
suitability (example, 6)
health education materials (example, 6)
dropouts (example, 7)
library use (example, 8)
nutrition surveillance (example, 9)
handicraft skills (example, 10)
nutrition status (example, 10)

Almost all of the words and phrases listed above can be understood by an educated person who knows the English language. An evaluator, however, can not deal with all words in their common meanings. For instance, in example 1 above, words "determine" and "relative" may be used in the dictionary meanings of the words, but not such words as "effectiveness", "new literates", "content" or "format". These words will have to be specially defined by the trainer-evaluator within the program context of his or her study.

Let us look at some of the other words and phrases listed above. What do we mean by "effectiveness"? Wouldn't "effectiveness" mean different things in each of the four examples (1, 2, 5 and 9 above); and be indicated by the appearance of different kinds of behaviors? It certainly would. What is a "new literate"? One who comes out of an adult literacy class or anyone who acquired his or her literacy recently? What level of literacy will be acceptable for a person to be called "new literate"? What "content" will be considered in the evaluation study - only the "content" of primers or of all reading materials? What aspects of "format" will be included in the study of instructional materials? What qualities will be examined in making judgments about "format"?

What is a "disabled person"? In the particular study referred to above only cripples were considered disabled. What is "social and economic rehabilitation"? Should the cripple be able to get married, become a village leader, and have his own house or hut to be considered "socially and economically rehabilitated"? What is a "dropout"? Is a "dropout" different from a "pushout" - the student who was asked to leave because he or she failed on the intermediate examination?

What constitutes "library use"? Would the borrowing of a prescribed textbook from the library constitute acceptable "library use"? What is "nutrition surveillance"? How is it different from "nutritional status"? What skills will be included in "handicraft skills"?

To sum, the trainer-evaluator in the elaboration and design of his or her evaluation study:

(1) will use some words in the dictionary meanings of the words (determine, relative);

(2) will use some words in their technical meanings as supplied by educators, psychologists and others (attitude, motivation, learning);

(3) will have to redefine some common words and also some technical words in terms of his or her special program intent and context (suitability, effectiveness, efficiency, rehabilitation, dropout, curricula); and

(4) will have to have standards for judging qualities of instructional products.

To do what is expected of trainers-evaluators, they must understand the concepts of: (a) concept analysis, (b) indicators of change, and (c) standards of quality of instructional products. We will deal with these in the following in the order given above.

Concept Analysis

Evaluation and ambiguity do not go together. One must know exactly what is meant by the evaluation question that is to be answered. This means that one must assign exact meanings to the concepts used in the evaluation question. For example, what is the meaning, precisely, of self-reliance, a development concept widely used in Tanzania? What precisely is the meaning of humanism which is the basis of the development philosophy in Zambia? What indeed is integrated rural development, rehabilitation of economically distressed families, teacher effectiveness, library use, administrative support, or a dropout?

The problem of conceptual analysis is the problem of definition; we want to determine the outline and limits of the words we use. We need to unpack concepts. We need to specify. We need to ask: What do I really mean? And we also need to ask: What do I not mean?

Philosophers have developed systematic methods of how to do concept analysis. Soltis[1] has suggested three different types of analyses that can be undertaken singly or in combination, as part of the definition of concepts and, subsequently, of indicators:

1. Generic type analysis. Generic is what is general and comprehensive and which pertains to every member of a class or category. In generic type analysis, therefore, we look for the general features of the concept being analyzed. Suppose we are analyzing the concept "development". As part of generic type analysis of development, we will ask the question: What general features must a social change program or activity have to be classified as development?

As we think of the "model" cases of development, we must also think of the contrary cases. For example, in the generic type analysis of development we should ask: Is economic growth in the South African apartheid regime an instance of development? If not, why not? The answers to this question should become part of our definition of development.

---

[1] Jonas F. Soltis, An Introduction to the Analysis of Educational Concepts (2nd edition). Reading, Mass.: Addison-Wesley Publishing Company, 1978.

Thus, we must work, on the one hand, with clear and standard cases to draw the necessary features of the concept. On the other hand, we must test our list of features for necessity and sufficiency by thinking of counterexamples. On the basis of this testing, we must modify, quality and reject some features to have a workable definition of our concept.

2. Differentiation type analysis. To differentiate is to distinguish, to understand the distinct and specialized character of something. Differentiation type analysis may be undertaken in lieu of generic type analysis or in addition to it, to tighten up your definition of concepts even further.

The prior question in differentiation type of analysis is - What are the different basic meanings of the concept being analyzed? We should search for the dominant standard uses of the concept (for instance, "development") by means of examples. We should then categorize uses into types - political development, educational development, cultural development, socio-economic development. Search for distinguishing marks which can be used to separate types should follow. After this we must test the typology (categorization or classification) developed with examples and counterexamples. The scheme of relationships that exists between various types classified may then be identified and a definition of the type in which we are interested must be elaborated.

3. Conditions type analysis. Finally, in conditions type analysis, we must ask the question about context conditions that govern the use of a

concept such as "development". What are the necessary conditions which must be present for development to happen? Are there contexts where the conditions hold but development does not occur? This might lead to a further search for conditions. In the end, the set of conditions established must be tested for necessity and sufficiency.

The three types of analyses involve mental moves on the chessboard of thought. Mastering these skills is not easy, but is by no means impossible. We learn to do such analyses by imitating and emulating; and by actual doing and testing. The skills can be developed by learning to criticize our own definitions and inviting friends and colleagues to find faults with our analyses of concepts.

Let us consider some examples:

Health education materials. The definitional problems here may be quite simple. We may decide to include only teacher made materials and exclude all commercially produced materials. We may decide to include posters, charts, and flannelgraphs but not "only print" materials such as leaflets or folders. Or we may include "only print" materials and none other.

Dropout. We may define as a dropout only that person who left the program on his or her own personal decision, not the one who was failed and asked to leave. We may not include one who was offered a position in the course but never joined. That is, dropouts may be separated from "no-shows". Again, a person who came back to join the course or the standard from which he or she had left, may be treated first as a "dropout" and later as "repeater".

Nutrition status. Nutrition status may be based only on the nutritional surveillance of pregnant mothers, lactating mothers, infants and children up to five years old. Older childrer may or may not be included. Some nutritional status reports may decide to cover everybody, including the very old.

Effectiveness. Effectiveness in one case (Example, 5 in the beginning of this chapter) may be defined as nothing more than scores on the annual English test set for Form IV students. In the case of the effectiveness of the role of chief back in their communities, the definition of effectiveness will have to be developed in terms of program objectives and the field context. The effectiveness may include administrative criteria, service criteria and political criteria. Is the community politically sympathetic to the ruling party? Do most people feel satisfied with the chief? Do they see him as a ruler or as a leader? Has the chief been able to establish new institutions such as cooperatives and primary courts in the community? etc., etc. In other words, the concept effectiveness will have to go through the process of conceptual analysis within that particular program and social context.

The questions we are asking in each case are: What are the meanings of the concept we are analyzing? What are not the meanings? Through what operations will the concept we are analyzing appear? This is what is often referred to as operationalization. In other words, the concept is defined in terms of operations that have concrete existence.

The basic purpose of concept analysis in the context of an evaluation study is to assign to the concept an invariance of meanings. Invariance means an absence of variation: thus, everyone using the concept should be able to assign the same meaning to the concept.

Indicators of Change

Often the problem of conceptual analysis is linked with a second problem - that of the development of indicators. After some abstract concepts such as effectiveness, efficiency, rehabilitation and motivation have been analyzed (unpacked and their different parts specified), another problem arises: How do we know that these abstract things actually exist in the field and are changing by some degree, in some direction?

The essential question in evaluating the impact of a development effort or the effectiveness of a development training program is - Did things change? Is "After" better than "Before", on the basis of some selected criteria? In other words, one must ask if a social system is in better shape after the application of some development strategy than it was before. Similarly, one must ask if the trainees of a development training program are better prepared to perform their tasks after their training than they were before.

The betterment in the condition of a society, or in the capacity of a development worker to undertake assigned tasks is not, however, always visible to the "naked eye". Things such as individual motivation and commitment, problem-solving capacity, political awareness, community cohesiveness, responsiveness of social institutions, and the quality of life, in general, are elusive concepts. We have to look for the signs, for some concrete manifestations of behavior which will indicate that a community has developed; or that the trainees of a training program are performing better. These signs are what we call indicators.

The process of developing indicators is complex, to say the least. Indicators must be valid, they must be concrete, and they must be parsimonious (that is the list of indicators for a condition should not be impractically long). To be able to engage in indicators research, one must understand the dynamics of psychological and social processes; must have sufficient grounding in logic and social science theory; and have extensive knowledge of human behavior. We can not, however, wait for ever to become experts. As practitioners and evaluators, we must learn to develop good enough indicators. Conceptual analysis as discussed in the earlier section is the necessary first step in the process of developing indicators of development or the effectiveness of development training.

An Introduction to Indicators Research

Indicators research has emerged as an important area of research in its own right over the last twenty years.

Economic indicators. Economic indicators have been the oldest and most frequently used. Most of us are familiar with the Gross National Product (GNP) per capita, the most widely used economic indicator. Interest rates and rates of inflation are other economic indicators. Since the 1960's, considerable interest has been focussed on the development of what are called social indicators.[1]

Social indicators. In the following pages, we present a few random examples of social indicators that have been developed and tested by social scientists in recent years, to give the readers a sense of the current activity in indicators research:

---

[1] Raymond A. Bauer, Social Indicators. Cambridge, Mass.: The M.I.T. Press, 1966.

Example I.1: Measuring economic well-being

The following components were identified that must be reflected in measuring the economic well-being of a family:

Cash income

Net worth of assets owned by a family

A family's endowment of human capital

The variability of income over time

Intrafamily transfers

The impact of government expenditures and taxes, and

Leisure and nonmarket productive activities.[1]

Health indicators as social indicators. Health indicators can be seen as a special case of social indicators. While most of the indicators shown in the example below relate to countries, they are transferable for use at the regional and community levels as well.

Example I.2: Health indicators with numerical values

A set of health indicators along with their values during the years 1975-77 are given below to show how enlightening such data can be:

Indicators	Kenya	Tanzania	Zambia	USSR	US
Life expectancy at birth	53	51	48	70	73
Infant mortality per thousand births	51	-	-	-	16
Population per physician	8,840	18,490	10,370	300	600
Percentage of population with access to safe water	17	39	42	-	-

---

[1] Marilyn Moon and Eugene Smolensky (Eds.), Improving Measures of Economic Well-Being. Institute for Research in Poverty Monograph Series. New York: N.Y.: Academic Press, 1977.

Daily per capita calorie supply as percentage of requirement	91	86	90	138	133
Adult literacy (which has been found to be a good health indicator)	40%	66%	39%	99%	99%
Health expenditures as percentage of GNP	1.8	1.9	2.6	2.7	3.3
Per capita public expenditures on health in US$	4	3	13	90	259

Science indicators as social indicators. Science indicators are also social indicators since they indicate the level of science and technology in a society.

Example I.3: Science indicators

To give the reader an idea of what kinds of indicators are used to develop profiles of levels of science and technology in a society, we give a few examples of science indicators. The list is not complete:

Research and development (R&D) expenditures as a percent of gross national product (GNP)

Patents awarded for original scientific inventions

Trade balance in technology-intensive products

Number of scientists produced by universities

Students in the science track in secondary schools, etc.[1]

Educational indicators as social indicators. Educational indicators are an important part of the total profile of a society.

---

[1] National Science Foundation, Science Indicators 1972. Washington, D.C.: United States Government Printing Office, 1973.

Example I.4: Categories of educational indicators

The following list of educational indicators was developed by Gooler:[1]

ACCESS

How many and what kinds of people participate in educational activities

Retention rates in educational activities

Catalog of existing/available educational activities or services

ASPIRATIONS

Description of needs and desires of various kinds of people

Individual self-assessments of personal capabilities

Description of institutional goals

ACHIEVEMENT

What people know, do, and feel

What people have earned (degrees, diplomas, certificates)

What is taught

IMPACT

Consequences of having schooling

Impact of education on social/economic/cultural systems

Consequences of not having schooling

RESOURCES

Capital, personnel, and material expenditures

Quality of human resources

---

[1] Dennis D. Gooler, "The Development and Use of Educational Indicators" in Educational Indicators: Monitoring the State of Education. Proceedings of the 1975 ETS Invitational Conference. Princeton, N.J.: Educational Testing Service, 1975, page 15.

Cost to benefit/effectiveness ratios

Quality of educational climate

Time

"Indicators of 'Indicators of Indicators'"

Quite often one cycle of indicators development may not be enough. We may have to develop 'indicators of indicators'. Yet, even that may be unsatisfactory and we may need to work on "indicators of 'indicators of indicators'". The following example should be pursued.

Example 1.5: Master social indicators

An American policy research institute[1] has suggested the following social indicators for developing a social report on the health of a nation:

Individual level	Societal level
Health	Health
Opportunity	Opportunity
Environment	Environment
Standards of living	Standards of living
Public safety	Public safety
Learning, science and culture	Learning, science and culture
Democratic values	Democratic values

Let us focus on "Democratic values" which through a concept analysis has been found to be an important part of the goodness of a society. (Note here the pervasive and interactive relationship between "concept analysis" and "indicator writing".)

---

[1] Educational Policy Research Center, Towards Master Social Indicators. Menlo Park, Ca.: Stanford Research Institute, 1969.

On pages 130 to 133, we have included their treatment of the concept of democratic values to demonstrate the conceptual analysis undertaken and the process of moving towards indicators and measures.

We should note how the concept of democratic values has been broken down into such dimensions as:

(A) At the individual level

1. Absence of severe threat
2. Freedom from undue social restrain, and
3. Freedom for personal growth

(B) At the societal level

1. No or minimal freedom
2. Adequate for social
3. Synergetic freedoms.

We should also note, how the "attainment categories" of democratic values in the two charts have been broken down into subcategories and then into possible indicators.

Indicators, we should note are the types of things we can see, hear, sense, judge and score.

Indicators of Interest to the Trainer-Evaluator

The work of indicator writing can not be undertaken lightly. The examples above from the literature on indicators research should help us understand the complexity and sophistication of the process.

Indicator writing, however, is not something that is done by supermen and is beyond the capacity of us trainers-evaluators. Indeed, in the everyday mundane settings of our work as trainers and evaluators, we will

Table 1

DEMOCRATIC VALUES (INDIVIDUAL)

Attainment Categories	Subcategories	Possible Indicators
Absence of severe threat	1. Freedom from severe restrictions on job, housing, schooling, public accommodation, voting	1. Measures of segregation
	2. Absence of strong stereotyping of minority by majority and vice versa	2. Attitude studies of whites toward blacks and vice versa
	3. No strongly patronizing or manipulative attitudes of majority towards minority	3. Measures of felt equality
	4. Safety from physical threat by majority	4. Incidence of vigilance-type activity, police brutality, etc.; subjective feeling of fear of this type of action
Freedom from undue social restraint	*1. Freedom of speech, press, assembly, privacy, due process	1. Number of law suits concerning these issues; newspaper space devoted to them; tenor of editorial comments
	*2. Freedom of religion, equality before law, antidiscrimination and antilibel protection	2. Percentage consulted on job decisions, feeling free to protest job decisions; participating in school discussions, feeling free to participate in school discussions
	*3. Open housing, public accommodations, choice of neighborhood, child-bearing	*3. Measures of alienation, anomie

Reproduced from Towards Master Social Indicators, Educational Policy Research Center, Stanford Research Institute, 1969, page 32.

Table 1 (Continued)

Attainment Categories	Subcategories	Possible Indicators
Freedom for personal growth	1. Trust of others; overall feeling of security	1. Content analysis of mass media
	2. Openness towards the views, beliefs, needs of others	2. Various public attitude polls
	3. Tolerance of diversity and of deviants	3. Various public attitude polls
	4. Responsiveness to others' views; willingness for others' views to be institutionalized	4. "Open-minded" newspaper treatment of all viewpoints; acceptance of radicals and conservatives by business
	5. Dignity accorded to those of other groups	5. Attitude polls

Table 2
DEMOCRATIC VALUES (SOCIETY)

Attainment Categories		Subcategories		Possible Indicators
No or minimal freedom	*1.	Abridgment of civil rights	*1.	Analysis of civil suits; desegregation
	*2.	Abridgment of legal rights	2.	Analysis of court decisions, arrest patterns
	*3.	Abridgment of human rights	*3.	Institutional discrimination in employment, housing, education, etc.
Adequate for social	*1.	Freedom of speech, press, assembly, privacy, due process	1.	Influence of public in decisions affecting it, voting patterns, false arrests, availability of free legal aid
	*2.	Freedom of religion, equality before law	2.	Private business, government discrimination measures
	3.	Freedom from libel, illegal monopoly, illegal business practices, discrimination	3.	Adherence to regulations, BBB activities, court records, data on actual minimum wages, trends in equal pay for equal work, trends in new laws
Synergistic freedoms	1.	Coincidence of institutional and individual goals	1.	Measures of insurrectionism, riots, civil disobedience, drop-outism, alienation, anomie, generation gaps
	2.	Attainment of full civil, legal, and human rights	2.	Number germane law cases

Reproduced from Towards Master Social Indicators, Educational Policy Research Center, Stanford Research Institute, 1969, page 39.

Table 2 (Continued)

Attainment Categories	Subcategories	Possible Indicators
	3. Full participative democracy	3. Degree to which one-man, one-vote doctrine holds; voting records, influence of machine politics, availability of candidates representing voter spectrum, degree of power elitism

have to develop our own indicators. These will have to be developed within the context of our own programs, for such qualities and conditions as:

Teacher effectiveness

Self-reliance

A successful extension agent

Cooperative behavior

Participative planning

Social and economic rehabilitation

Positive developmental impact.

It is not possible to actually work out the indicators for each of the randomly chosen concepts and constructs listed above. That will be a task of substantial dimensions outside the scope of this short chapter. Even if we did include indicators of teacher effectiveness, self-reliance, change agent's success, cooperative behavior, participative planning, social and economic rehabilitation and positive developmental impact, we will leave much out. It would be impossible to include all the indicator writing needs of trainers-evaluators in all the various development sectors.

More importantly, however, standardized sets of indicators are not possible because indicators writing has to be often done in terms of a particular content; and relative to the concept definitions developed in a particular program setting. The general principle is to go from a first abstraction through categories and subcategories to behavioral manifestations: to things which can be seen, heard, touched, sensed, judged and scored.

Validating indicators. Unfortunately no standard formulas can be suggested for writing indicators and testing their goodness - their reliability and their validity. Ultimately the goodness of indicators will be proven through their testing-by-use.

It would always be a good idea, however, for trainers-evaluators to pre-test their indicators through peer reviews. They should show their indicators and their 'indicators of indicators' to their colleagues and let them critique their work.

The indicators-instrument connection. We will be discussing evaluation instruments and their construction and use in the next chapter. However, we wish to point out here the clear and direct connection between the process of indicator development and the process of instrumentation. In the construction of tools and instruments we merely take the next logical step from indicator development. We ask: What data or evidence should be collected to demonstrate the existence of or change in the indicator-related behavior or condition? How to collect or elicit the required data or evidence? What aids (tests, tapes, questionnaires, schedules, etc.) might be used for recording the data or evidence? (See the Figure on next page.)

Standards of Quality of Instructional Materials

Trainers-evaluators will also have to face the problems of developing instructional norms (What is a good training method?); and standards of quality to judge a variety of instructional materials (Is this a good poster?).

Program Objectives →	Program Actions →	Program Effects and Where the Effects Appear →	Indicators of Effects →	Items for Instruments
O_1	A_1			
		Individuals	i_1 i_5	Test-Items
		I_1 I_4	i_2 i_6	t_1, t_2, t_3, t_4, ...
O_2	A_2	I_2 I_{n-1}	i_3 i_{n-1}	t_{n-1}, t_n
		I_3 I_n	i_4 i_n	
		Groups	g_1 g_5	Questionnaire Items
		G_1 G_4	g_2 g_6	q_1, q_2, q_3, q_4, ...
O_3	A_3	G_2 G_{n-1}	g_3 g_{n-1}	q_{n-1}, q_n
		G_3 G_n	g_4 g_n	
		Organizations	o_1 o_5	Interview Schedule Items
		IS_1 IS_4	o_2 o_6	v_1, v_2, v_3, v_4, ...
O_{n-1}	A_{n-1}	IS_2 IS_{n-1}	o_3 o_{n-1}	v_{n-1}, v_n
		IS_3 IS_n	o_4 o_n	
		Communities	c_1 c_5	Observation Schedule Items
		C_1 C_4	c_2 c_6	b_1, b_2, b_3, b_4, ...
O_n	A_n	C_2 C_{n-1}	c_3 c_{n-1}	b_{n-1}, b_n
		C_3 C_n	c_4 c_n	

Figure 1. The flow from program objectives through indicators to instrumentation.

The question of general instructional norms has already received attention in our discussions of concept analysis and indicators development above. The problem of developing standards of quality to judge instructional products such as posters, charts, films, primers and books; or to judge overall instructional facilities and settings is a special one and deserves separate attention.

Here we are faced with judging the information, design, technology and aesthetic attributes of instructional items.

Trainers-evaluators will have to learn a lot about instructional systems technology to be able to develop these standards for the evaluation of instructional products. Some "Appraisal checklists" on various items of instructional materials developed by Heinich, Molenda and Russell[1] have been included in the following pages. Special attention should be paid to the section on "Rating".

---

[1] R. Heinich, M. Molenda and J. Russell, Instructional Media. New York: John Wiley and Sons, 1982.

APPRAISAL CHECKLIST: GRAPHIC MATERIALS

Title or Content of Graphic:____________________________

Producer/Distributor (if known)____________________________

Series (if applicable)____________________________

Date (if known)____________________________

Format
- ☐ Drawing
- ☐ Chart
- ☐ Graph
- ☐ Poster
- ☐ Cartoon

Objectives (stated or implied):

Brief description

Entry capabilities required

Prior subject matter knowledge
Visual skills
Others

RATING	High		Medium		Low
Simplicity (few elements or ideas to catch and hold the attention of the viewer)	☐	☐	☐	☐	☐
One main idea (to provide unity)	☐	☐	☐	☐	☐
Relevance to curricular objectives	☐	☐	☐	☐	☐
Color (attracts and holds attention)	☐	☐	☐	☐	☐
Verbal information (reinforces the ideas presented in the visual)	☐	☐	☐	☐	☐
Learner comprehension	☐	☐	☐	☐	☐
Legibility for classroom use	☐	☐	☐	☐	☐

Strong points:

Weak Points:

Recommended Action____________________________

Reviewer__________________

Position__________________

Date__________________

John Wiley & Sons, Inc., grants the reader permission to reproduce this Checklist for personal use only (Heinich, Molenda, and Russell. INSTRUCTIONAL MEDIA. 1982)

of print media — newspapers, periodicals, textbooks, etc — and range from comic strips intended primarily to entertain to drawings intended to make important social or political comments. Humor and satire are mainstays of the cartoonist's skill.

Cartoons are easily and quickly read and appeal to children and adults alike. The best of them contain wisdom as well as wit. As such, they can often be used by

BEST COPY AVAILABLE

APPRAISAL CHECKLIST: AUDIO MATERIALS

Title ____________________

Producer/distributor ____________________

Series (if applicable) ____________________

Date (if known) __________ Price __________

Format	Speed	Time
__Record	__rpm	__min
__Reel-to-reel	__ips	
__Cassette		

Objectives (stated or implied):

Brief description:

Entry capabilities required:

-Prior subject matter knowledge
-Audio skills
-Other

RATING	High		Medium		Low
Accuracy	☐	☐	☐	☐	☐
Sound quality	☐	☐	☐	☐	☐
Student involvement	☐	☐	☐	☐	☐
Interest level	☐	☐	☐	☐	☐
Vocabulary level	☐	☐	☐	☐	☐
Overall value	☐	☐	☐	☐	☐

Strong points:

Weak points:

Reviewer ____________________

Position ____________________

Recommended action ____________________

Date ____________________

John Wiley & Sons, Inc., grants the reader permission to reproduce this Checklist for personal use only (Heinich, Molenda, and Russell: INSTRUCTIONAL MEDIA, 1982)

BEST COPY AVAILABLE

APPRAISAL CHECKLIST: FILM

Title____ Format __16 mm
Producer/distributor____________________________________ __8 mm
Date (if known)__________ Length__________ __other
Audience/grade level____________________________________ ______
Subject area(s)______________________________________ __color
__black/white

Objectives (stated or implied):

Brief description (include presentation style: animated, dramatic, etc.):

Entry capabilities required:

--language ability
--prior subject matter knowledge

RATING	High		Medium		Low
Likely to arouse student interest	☐	☐	☐	☐	☐
Technical quality	☐	☐	☐	☐	☐
Opportunity for viewer participation	☐	☐	☐	☐	☐
Relevance to curriculum (or learning task)	☐	☐	☐	☐	☐
Accuracy of information	☐	☐	☐	☐	☐
Scope of content	☐	☐	☐	☐	☐
Organization of content	☐	☐	☐	☐	☐
Student comprehension	☐	☐	☐	☐	☐

Strong points:

Weak points:

Reviewer________________
Position________________
Recommended action________________________ Date________________

John Wiley & Sons, Inc., grants the reader permission to reproduce this Checklist for personal use only (Heinich, Molenda, and Russell. INSTRUCTIONAL MEDIA 1982)

BEST COPY AVAILABLE

APPRAISAL CHECKLIST: TELEVISION

Series title ______________________________

Program title (or number) ______________________________

Producer/distributo ______________________________

Production date ______________________ Program length ______ min.

Intended audience/grade level ________ Subject area ______________________

Objectives (stated or implied):

Brief description:

Entry capabilities required:

—prior knowledge
—reading ability/vocabulary
—math ability

RATING	High		Medium		Low
Likely to arouse student interest	☐	☐	☐	☐	☐
Technical quality	☐	☐	☐	☐	☐
Provides meaningful viewer participation	☐	☐	☐	☐	☐
Objectives relevant to curricular needs	☐	☐	☐	☐	☐
Focuses clearly on objectives	☐	☐	☐	☐	☐
Evidence of effectiveness (e.g., field-test results)	☐	☐	☐	☐	☐
Teacher's role clearly indicated	☐	☐	☐	☐	☐
Provides guide for discussion/follow up	☐	☐	☐	☐	☐

Strong points:

Reviewer ______________________

Weak points:

Position ______________________

Recommended action ______________________________ Date ______________________

John Wiley & Sons, Inc., grants the reader permission to reproduce this Checklist for personal use only (Heinich, Molenda, and Russell INSTRUCTIONAL MEDIA, 1982)

BEST COPY AVAILABLE

APPRAISAL CHECKLIST: SIMULATION/GAME

Title ________________________

Publisher/distributor ________________________

Publication date ____________

Number of players ____________

Special equipment or facilities needed.

Intended audience/grade level ____________

☐ Has gaming features (competition, scoring)

☐ Has simulation features (role playing)

Playing time ____________

Subject area ________________

Objectives (stated or implied):

Brief description:

Entry capabilities required

- —prior knowledge
- —reading ability
- —math ability

RATING	High		Medium		Low
Likely to arouse student interest?	☐	☐	☐	☐	☐
Players practice meaningful skills	☐	☐	☐	☐	☐
(Game) Winning dependent on player actions (vs. chance)	☐	☐	☐	☐	☐
(Simulation)Validity of game model (realistic, accurate)	☐	☐	☐	☐	☐
Technical quality (durability, attractiveness, etc.)	☐	☐	☐	☐	☐
Evidence of effectiveness (e.g. field-test results)	☐	☐	☐	☐	☐
Clear directions for conducting game?	☐	☐	☐	☐	☐
Players' instructions clear and concise?	☐	☐	☐	☐	☐
Debriefing guide included?	☐	☐	☐	☐	☐

Strong points:

Weak points:

Recommended action ________________________

Reviewer ________________

Position ________________

Date ________________

John Wiley & Sons, Inc. grants the reader permission to reproduce this Checklist for personal use only (Heinich, Molenda, and Russell, INSTRUCTIONAL MEDIA, 1982)

BEST COPY AVAILABLE

Summary

To go from the stage of having an "evaluation question" to the development of tools for collecting "evaluation data", trainers-evaluators must cross the two important bridges of concept analysis and indicator writing. Concept analysis involves three interrelated analytical strategies: generic type analysis, differentiation type analysis and conditions type analysis. The basic questions are: What does the concept mean? What does the concept not mean? What operations lead to the emergence in real-life of the concept and under what conditions?

Indicator writing is an extension of the idea of concept analysis. The questions now asked are: What concrete entities and behavioral data (things we can see, hear, touch, sense, judge and score) will be considered as evidence that the concept in question exists in real-life; or is changing in some measure in some direction? The continuous link from program objectives, through program actions, program effects and where those effects appear, indicators of effects to items for different tools and instruments is indicated. The special need for standards of quality for judging instructional materials has been discussed.

Things to Do or Think About

1 Examine the definitions of some terms given below.

(a) Which ones are definitions of common words that you are ready to accept as universal?

(b) Which ones are definitions of technical words that, again, you will accept as generally applicable?

(c) Which ones are definitions of technical words but which must be modified through the process of concept analysis within your particular program and social context?

Truth: Conformity to fact or reality; to rule, standard, model, pattern or ideal. Faithfulness to the facts of nature, history or life.

Learning: Permanent change in behavior that is the result of past experience either produced incidentally or through institutional learning through teaching.

Primary education: Those years of study during which no differentiation is introduced either in the forms of optional subjects or in the streaming of pupils towards different types of institution or education.

Dropout: Person who leaves school or college before completing the course of study.

2. Undertake the concept analysis of the concept of "self-reliance" in the Tanzanian setting.

3. Develop the indicators of "training effectiveness" for a hypothetical training course for rural health workers.

4. What would you want to see in:

a good primer
a good rural newspaper
a good poster or
a good teaching film?

CHAPTER X

EVALUATION TOOLS AND INSTRUMENTS

In Chapter I (What is Evaluation), it was pointed out that while "evaluation" and "research" differ significantly from each other in regard to their objectives and functions, the evaluator and researcher do share their general approaches, instruments and tools for data collection.

Later, in our discussion of the two basic paradigms of evaluation and research (Chapter VIII: Models of Evaluation), we indicated that the scientific paradigm and the naturalistic paradigm are rooted in two different worldviews and demand two quite distinct methodological approaches to studying the world and making knowledgeable statements about this world. Consequently, the kinds of instruments chosen by evaluaters using the scientific paradigm are different from those chosen by evaluators using the naturalistic paradigm. The scientific paradigm typically demands structured instruments of data collection. Within the naturalistic paradigm, the evaluator himself or herself is the primary instrument: the approaches, the tools and the instruments in use are unstructured. The following display should clarify the choices of tools and instruments made by evaluators and researchers within the two paradigms - the scientific and the naturalistic:

	Research	Evaluation
Scientific Paradigm	Structured tools and instruments Tests Questionnaires Interview Schedules Observation Schedules/ Checklists	Structured tools and instruments Tests Questionnaires Interview Schedules Observation Schedules/ Checklists

	Tools for Content Analysis of Records/ Documents	Tools for Content Analysis of Records/ Documents
Naturalistic	Unstructured tools and instruments In-depth Interviews Unstructured Observations Content Analysis of Records/Documents	Unstructured tools and instruments In-depth Interviews Unstructured Observations Content Analysis of Records/Documents

In the following, we will deal with the various tools and instruments recorded in the display above.

Scales of Measurement

Measurement is an important part of evaluation. We often need to go beyond the crude comparisons of "good-better-best" and "big-bigger-biggest." To do this we need standard yardsticks with which we can take the measures we want; and can state *how much* of a difference exists between two entities, and in what direction.

Unfortunately, in the social sciences we do not have the benefit of such tools as micrometers, carbon dating and atomic clocks. Our measures and yardsticks are often quite crude. We need to understand, however, the nature of scales that are available to us; and we need to understand their possibilities and their limitations.

The nominal scale. The nominal scale does not really measure, it only nominates objects to categories. The classification of adults in a community into males and females, and assigning them numbers (1 for males, 2 for females), will be an example of using a nominal scale.

We need to understand that numbers used in the nominal scale mean nothing about the value of categories, except to show that they are different. In the above example, 2 (for females) is not twice of 1 (for males). The numbers 2 and 1, in this particular context, can not be added or subtracted from each other in any meaningful way.

The ordinal scale. The ordinal scale introduces ordering to the nominal scale. The categories can now be ranked in an order of succession as "First, Second and Third," or "Good, Medium and Poor."

The ordinal scale, again, could be assigned numerical values: for example, 5 for Good, 3 for Medium and 1 for Poor. But, once again, 5 is not five times 1 in terms of the scale, nor is 6 (two mediums) better than 5 (one good).

The interval scale. The interval scale, as the name suggests, has intervals which make mathematical sense. On a meter rod, the difference between 3 and 5 centimeters is the same as the distance between 53 and 55 centimeters.

Scores on an achievement test are in reality ordinal data, but we can often treat it as if it was interval data. We can say that B made twenty points (or twenty intervals) more than A. However, if B had made 40 points and A had made 20, we could not say that B is twice as good as A. To be able to make that kind of statement, we will need ratio scales.

The ratio scale. The ratio scale, in addition to being an interval scale, has an absolute zero. This means that 25 is 5 points more than 20, and that 60 is three times as good as 20. Thus, the ratio scale permits us to work out ratios and proportions. Two meters is twice as long as 1 meter. Something can be twice as hot as another.

We need to keep the properties of various scales of measurement in view as we deal with data from our various evaluation tests, tools and instruments.

I

Tests

Anyone who has been to school has been subjected to tests (or exams as they are popularly called.) Tests are a usual tool of the evaluator working within the scientific paradigm. Evaluators working within the naturalistic paradigm may avoid formal testing, but they can not, however, avoid judging and selecting. In judging and selecting, naturalistic evaluators will perhaps use observation of performance-related behavior within the relevant social context instead of formal tests.

Tests, or achievement tests as they are often called, are tests of knowledge, skills and performance. Tests may be made to measure knowledge in arithmetic, biology, nutrition or animal husbandry; research skills, diagnostic skills or graphic skills; or actual performance in a role.

Tests can also be used to measure aptitudes (natural or acquired abilities or bents of mind). In fact an aptitude test can be seen to be a special kind of achievement test.

Evaluators may sometimes be interested in testing attitudes (value dispositions and opinions). Attitude testing will be discussed later as part of questionnaires and interviews.

Having gone through lot of achievement testing in our lives; and, perhaps, having ourselves written and administered tests as teachers and trainers, we might think of tests as relatively simple to make, to administer and to interpret. This is not really true. There are many complexities involved as the discussion that follows should show.

Standardized Norm-Referenced Tests and Criterion-Referenced Tests

Tests may be made for one particular group (community health workers under training in a special workshop) or for a large regional or national population (all VIII grade students in Kenyan schools or even East African schools).

In the first case, the test will most likely be designed to measure whether the community health workers have learned most of what they were taught within the special workshop that they had attended. The criterion of success may be a score of at least 80 out of the possible 100 on an achievement test specially designed for that group. This would be an example of a criterion-referenced test.

In the second of the two cases above, the test will most likely be designed to measure how well a student, a class, or a school is doing in comparison to other students, classes and schools tested on the same test of VIII grade mathematics or English or civics. To be able to make those comparisons, we will have to have norms - how is an average VIII grade student supposed to perform on this particular test. When these norms do become available, the test becomes norm-referenced and standardized.

The process of standardization of tests for development of norms is itself quite standardized now. We do not discuss it here because trainers-evaluators will most often be dealing with criterion-referenced tests. Those are the tests we will focus upon in the following discussion.

Training Objectives and Testing Objectives

Training and testing objectives should match with each other. It would be patently unfair to test trainees on things they were never taught. This means that the test writer should have available to him or her a clear and detailed statement of the instructional objectives of a training course, to be able to make a test that will measure the impact of the course, effectively.

Professors Benjamin S. Bloom, D. R. Krathwohl and their associates have developed taxonomies of instructional objectives that should interest both trainers and test makers. The basic outlines of their taxonomies are reproduced below:

INSTRUCTIONAL OBJECTIVES IN THE COGNITIVE DOMAIN[1]

- 1.00 Knowledge
  - 1.10 Knowledge of specifics
  - 1.11 Knowledge of terminology
  - 1.12 Knowledge of specific facts
  - 1.20 Knowledge of ways and means of dealing with specifics
  - 1.21 Knowledge of conventions
  - 1.22 Knowledge of sequences
  - 1.23 Knowledge of classifications and categories
  - 1.24 Knowledge of criteria
  - 1.25 Knowledge of methodology
  - 1.30 Knowledge of the universals and abstractions in a field
  - 1.31 Knowledge of principles and generalizations
  - 1.32 Knowledge of theories and structures

- 2.00 Comprehension
  - 2.10 Translation
  - 2.20 Interpretation
  - 2.30 Extrapolation

- 3.00 Application

---

[1] B. S. Bloom et al., Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: David McKay, 1956.

4.00 Analysis
4.10 Analysis of elements
4.20 Analysis of relationships
4.30 Analysis of organizational principles

5.00 Synthesis
5.10 Production of a unique communication
5.20 Production of a plan, or proposed set of operations
5.30 Derivation of a set of abstract relations

6.00 Evaluation
6.10 Judgements in terms of internal evidence
6.20 Judgements in terms of external criteria

INSTRUCTIONAL OBJECTIVES IN THE AFFECTIVE DOMAIN[1]

1.00 Receiving (attending)
1.1 Awareness
1.2 Willingness to receive
1.3 Controlled or selected attention

2.00 Responding
2.1 Acquiescence in responding
2.2 Willingness to respond
2.3 Satisfaction in response

3.00 Valuing
3.1 Acceptance of a value
3.2 Preference for a value
3.3 Commitment

4.00 Organization
4.1 Conceptualizing a value
4.2 Organizing a value system

5.00 Characterization by a value or value complex
5.1 Generalized set
5.2 Characterization

The test writer should not confuse the cognitive with the affective, or the ability to synthesize with the simple knowledge of universals and abstractions. We should realize that learning of information does not ensure real comprehension; and comprehension does not automatically lead to the ability

---

[2] D. R. Krathwohl, *et al.*, *Taxonomy of Educational Objectives, Handbook II: Affective Domain*. New York: David McKay, 1964.

to apply, analyze and judge. Similarly, it is possible to be positive verbally to particular entity or a position without genuine commitment; and to have a set of discrete values that do not add up to a systematic and organized value system. We simply can not teach one thing and test for another.

Choosing the Test Content

It is obvious that one can not test everything that has been taught. One will have to take a small sample of all the knowledge taught, to be included in a test.

The sample of knowledge to be included in a test should be developed systematically from a detailed and comprehensive description of the subject matter taught. The two taxonomies presented above should be used for the description of subject matter taught: What factual knowledge was taught? What general principles and generalizations were communicated? What diagnostic skills and abilities to apply and transfer to other situations were underlined? What change in attitudes and values was reinforced as part of the resocialization for the new role?

Based on this comprehensive description, a sample of knowledge and values should be selected for test making.

Types of Test Items

A variety of test items can be written to be included in an achievement test.

True/False. A statement is written and the respondent is asked to check it as true or false.

Example.

Groundnuts and vegetables are
body-building foods. T/F

(Answer: True)

True/False items are comparatively easy to write. These are, however, of limited use in testing for depth of understanding. The advantage of easy scoring is balanced by a disadvantage. Respondents feel encouraged to guess answers when they do not really know the answer. As they make guesses, they have a 50:50 chance of being right.

Short answer and completion items. As the name suggests, these items require a short one- or two-word answer or the filling in of a blank.

Examples.

What do spittle and rubbish breed?

(Answer: Microbes.)

$$\begin{array}{r} 239 \\ \underline{-143} \\ = \end{array}$$

(Answer: 96)

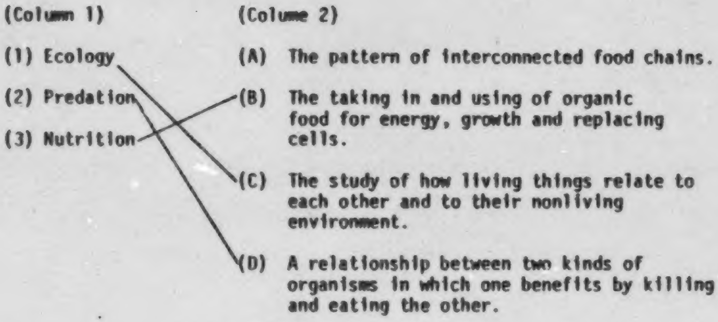
The manometer of the sprayer shows that it
has ______________.

(Answer: Pressure)

Short answer and completion items have to be written carefully so that more than one interpretation of the question/incomplete sentence is not possible. The wording of the item should elicit the information specifically required.

Matching. Matching involves pairing of items from two different sets or columns because of their similarity or correspondence according to some rule or relationship.

Example.



(Column 1)

(1) Ecology

(2) Predation

(3) Nutrition

(Colume 2)

(A) The pattern of interconnected food chains.

(B) The taking in and using of organic food for energy, growth and replacing cells.

(C) The study of how living things relate to each other and to their nonliving environment.

(D) A relationship between two kinds of organisms in which one benefits by killing and eating the other.

Matching items should be kept relatively short. Note that there are three choices under Column 1 and four choices under Column 2. This insures that matching will involve deliberate choices in all cases under Column 1.

If a choice under one of the columns is usable more than once, make that information available to students as a part of the question.

Multiple-choice. Multiple-choice items are the most versatile and effective form of test items. A multiple-choice item has a stem, followed by multiple options from which one or more could be selected.

Example.

A farmer should do early weeding of his cotton crop: [Stem]

(a) So that the cotton is not chocked

(b) So that weeds do not consume up the plants' food

(c) So that cotton gets enough air

(d) So that cotton has access to light

(e) So that cotton gets enough water

(f) Because weeds could breed dangerous insects for cotton

(g) To allow better growth of cotton

(h) To get a good cotton yield

[Options]

Most of the options above are correct. Choosing the right options and leaving out the incorrect ones will be like writing a short essay on the advantages of early weeding of the cotton crop.

Typically, multiple-choice items have no more than four or five options, unlike the item above which has eight options.

Essay. This is the easiest type of test to write and the most difficult one to score. When essay questions are carefully written, specifying exactly what is required, essay questions do provide the students with opportunities to analyze, synthesize and evaluate subject matter content. Objectivity of scoring of essay type questions can be increased if teachers themselves write model answers to their own essay type questions and then judge student responses according to the model answers.

Simultations. Simultations of various kinds provide exciting teaching and testing possibilities. Various types of "In-Tray/Out-Tray" simultations can be designed to test the performance abilities of trainees in life-like decision-making situations.

Pre-testing Tests for Improvement

Good test items have to test what they are supposed to test; and should be well-written so that they communicate the same one meaning to all readers clearly and unambiguously.

Item writing takes time, patience and skills. With time and patience, skills can be developed. One thing that test writers must do is to pre-test their tests; and go through careful revisions of their tests on the basis of their pre-testing.

After a more wide-scale use of a test in an evaluation study, the test should be revised again. Even if you will never use it again, the revisions will train you for better test writing for future evaluation studies.

Time Tests, Power Tests and Other Considerations in Administering Tests

Tests should be administered so as not to make respondents afraid and anxious - what is called "test anxiety" can become a serious problem. Indeed, within developmental settings, where we deal with adults (and also with government functionaries), we may find that we want to give a test but the adults concerned do not want to take the test. Sometimes a few test items may have to be hidden in an opinion questionnaire or an interview schedule.

When administering a test, the respondents should be comfortably seated and instructions in regard to how to complete the test should be fully explained.

Finally, tests can be time tests or power tests. Time tests have to be completed within a particular period of time: 45 minutes or an hour, for instance. At the end of this time, test papers are collected whether or not these have been completed. Power tests are given to determine how much the respondents have learned (and not how fast they can answer questions). In a power test, there are many more test items than there are in a typical time test, and time is allotted generously to students for completion of the test.

II

Questionnaires

Questionnaires are a set of questions, systematically arranged on paper to be asked from a group of respondents.

Spaces for recording answers are provided within the questionnaire. Questions may be open-ended, which means that the respondent will formulate the answer in his or her own words. Or, the answers may be structured, requiring the respondent to make multiple choices or to check on scales built into the questionnaire.

Questionnaires are often distributed by hand or sent by mail. When dealing with illiterate respondents or to ensure high rate of response, the evaluator may administer questionnaires in person. In such a case the questionnaire becomes a "structured interview."

Questionnaires should be short and well-designed. Instructions should be clear and easy to understand. A short introduction should provide the purpose of the questionnaire and explain how the data provided by the respondent will help the respondent or others in the community. Anonymity of respondents should be ensured.

In the preceding section on tests, we have suggested that tests are test of knowledge. Questionnaires also test knowledge but it is the "particular knowledge" of a person that questionnaires seek to find out about. It is not the general knowledge of subject matter but the private knowledge of a person - information personally available, his or her perceptions, attitudes and opinions of various kinds. (As we have indicated in the preceding section, some achievement test items may sometimes be hidden in a questionnaire.)

Local Adaptation of Available Questionnaires

Questionnaires can not be standardized (as norm-referenced tests). Trainers-evaluators will be making their own questionnaires to suit the special social and program context of their evaluation studies. However, questionnaires on similar subjects developed by other trainers-evaluators elsewhere may be adapted to fulfill particular needs. Many questionnaire items may be possible to lift and use with very little rewriting.

Types of Questionnaire Items

All the types of items that can be used in an achievement test, can also be used in a questionnaire: true/false, short answer or completion, matching, multiple-choice, short essay and even simulations. The only difference is that we are now eliciting "private knowledge," perceptions, attitudes, opinions and judgments. (Refer to the discussion on "Types of Test Items" under Section I above.)

Making Scales for Recording Attitudes and Opinions

In designing questionnaires which include lots of attitudinal and opinion questions, the evaluator will have to design scales to record the subject's responses.

Sometimes these scales may be in the simple form of

Yes Undecided No

or

Agree Don't know Disagree

To make the scale more sensitive, additional ordinates may be added, as

Strongly agree	Agree	Don't know	Disagree	Strongly disagree

Robert B. Jones[1] made an opinion survey of military officers in regard to what they thought of their "military assistance training courses" and used a scale with many dimensions. (The scale below also shows, in percentages, the responses actually received.)

MILITARY ASSISTANCE COURSES IN GENERAL
(Based on the Courses You Have Taken)

Creative	3.7	22.2	14.8	14.8	14.8	14.8	14.8	Uncreative
Organized	26.9	34.6	19.2	3.8	--	15.4	--	Disorganized
Hard	--	3.8	26.9	26.9	15.4	19.2	7.7	Easy
Relevant	3.8	34.6	26.9	19.2	7.7	3.8	3.8	Irrelevant
Flexible	--	11.5	26.9	15.4	11.5	23.1	11.5	Inflexible
Practical	7.7	30.8	26.9	19.2	11.5	--	3.8	Impractical
Exciting	--	11.5	30.8	19.2	19.2	11.5	7.7	Dull
Strong	3.8	15.4	34.6	23.1	7.7	7.7	7.7	Weak
Active	3.8	19.2	34.6	19.2	15.4	7.7	--	Passive
Demanding	3.8	7.7	26.9	19.2	15.4	19.2	7.7	Undemanding
Scientific	--	15.4	19.2	38.5	11.5	7.7	7.7	Artistic
Involving	3.8	15.4	38.5	19.2	11.5	3.8	7.7	Alienating
Objective	3.8	11.5	34.6	19.2	15.4	11.5	3.8	Subjective

---

[1] Robert B. Jones, The Influentials in Uniform: A Study of the Training, Attitudes and Role Perceptions of the United States Military Assistance Officers as Change Agents in Developing Countries. Doctoral Dissertation, Indiana University, 1974.

Modifiable	7.7	11.5	26.9	30.8	7.7	11.5	3.8	Unmodifiable
Motivating	3.8	19.2	26.9	34.5	7.7	--	7.7	Alienating

This scale is multi-dimensional since it uses many bipolar dimensions: creative-uncreative, organized-disorganized, hard-easy, etc. Also, the scale introduces seven ordinal points along each dimension of the scale.

Jones has collated data in percentages: What percentage of respondents chose each ordinal point on each of the dimensions of the scale? As we will discuss later, ordinates can be given numerical values as follows:

Creative						Uncreative
+3	+2	+1	0	-1	-2	-3

Such scores can then be aggregated. One has to learn to be careful about assigning numerical values or weights to responses on such scales, however.

Writing Good Questionnaires

Good questionnaires are made with clear objectives in view. They ask what the trainer-evaluater needs to know, avoiding unnecessary questions. But the important questions are not forgotten. Standard demographic information such as sex, age, occupation, income, etc. is always asked to be able to interpret responses received.

Item writing for questionnaires offers an additional set of problems since (1) they may ask for private knowledge that the respondents may be unwilling to part with; and (2) they elicit opinions and attitudes that the respondents may be unprepared to share honestly. Attitudes in regard to family planning, inter-marriages between people from different tribes, and taboo foods may not be honestly expressed. The respondent may supply "socially acceptable" responses. They may tell the trainers-evaluators what they want to hear rather than what the respondents actually think.

To solve some of these problems, writers of questionnaires may make the intent of an item less direct and may ask the same question in different ways within the same one questionnaire.

Once again, pre-testing of questionnaires is important before administering them on a large scale as part of an evaluation study. Such pre-testing will bring out many problems in the questionnaire.

Let us look at some examples:

Items	Comments
A district officer is asked: After information has been communicated to the chiefs/assistant chiefs in your area, how is this acted upon?	Can the district officer really tell? Wouldn't it be better to get this information from the chiefs themselves? Aren't we asking the wrong respondent?
A community level nutrition worker is asked: What do you engage in during your home visits?	Isn't this too general a question?
A subject is asked: Do you attribute your friend's failure to laziness?	What is laziness? Do we all mean the same thing by the word laziness? This is ambiguous.
A subject is asked: Do you think you were in good health during the period of the training course?	Do the subject and the evaluator understand the same thing by health? What, if the student has not been too well, but never too sick to miss classes for long? Shouldn't we ask the question in terms of days missed because of sickness? The question is ambiguous in its present form.
An extension worker under training is asked: Was your visit to the farmer useful?	Useful to whom? In what way? On the basis of what kind of evidence, using what criteria? Ambiguous, again.

The headmaster of the school is asked to judge the student-teacher's commitment to work in terms of: -- unsatisfactory -- below average -- average -- above average -- outstanding	How to ensure that the evaluator and the headmaster mean the same thing by the word commitment? Do we define commitment in terms of punctuality, or carrying an overload of work, or offering tutorials to weak students? How will the headmaster come to acquire the knowledge on which to base his/her judgments?
A local extension worker is asked by the evaluator: Are locally made audio-visual materials better than those produced elsewhere?	Does "elsewhere" mean in another locality? National headquarters? A commercial producer? Does better mean better in production values or in terms of instructional relevance? There are many ambiguities.
A cooperative assistant at the community level is asked by the evaluator: Are you highly motivated to make the best contribution to your work?	Wouldn't most of them say "Yes!"? Isn't it a loaded question?
An adult education officer is asked: How many of your earlier students still practice reading skills?	Is this question answerable?

Many of these problems may be caught in the process of pre-testing of the questionnaire. With practice, item writing for questionnaires will surely improve.

III

Interviews

Interviews are used by evaluators both for scientific inquiry and naturalistic inquiry. In the context of the scientific paradigm, interviews are structured or semi-structured (basically a structured interview, with some probing questions introduced to seek further explanations). In the context of the naturalistic paradigm, interviews are unstructured and are conducted in depth for developing "thick descriptions."

As we have indicated before, structured questionnaires when administered in person become structured interviews. The structured interview, therefore, has the same problems and concerns of design, item writing and display of items as does the structured questionnaire. (See Section II above.)

But since interviews are conducted in face-to-face situations, they pose some additional problems and challenges. The interviewee must be motivated to give the interview and to invest the time required for completing the interview. The interviewer should be able to establish trust and rapport without influencing the responses of the interviewee. In rural settings of developing countries, it may not be possible always to take the interviewee (especially a female interviewee) aside for a long private conversation. The interviewer should assure that an individual interview with a young mother does not become a family interview.

Sometimes family interviews may be just the thing we want. But then we should plan and work for a family interview. The point is that an individual interview should not get confused with a family or group interview.

It is also possible to use more than one interviewer in conducting an interview. A chief may be interviewed about his work by a full panel of interviewers.

In-depth Interviews for Naturalistic Evaluations

Interviews conducted as part of naturalistic evaluations are not structured. The interviewer starts not with questions but with themes. After establishing the general boundaries of the subject of the interview, the naturalistic interviewer lets the interviewee take over. The interviewer listens, with interest and sympathy, encouraging the interviewee to go on, to explain further, to come back to the point, to choose, to judge and to take positions. Such interviews typically take several hours.

In-depth interviews may be recorded by the interviewer in his diary in the presence of the interviewee or immediately after the interview when the interview is still fresh in the interviewer's mind. Tape-recording the interview may be preferable, if the interviewee willingly permits.

IV

Field Observation

Field observation, again, is a data collection strategy that can be used both within the scientific and the naturalistic paradigms. Once again, field observation within the scientific paradigm may be sampled and highly structured. Within the naturalistic tradition, field observation will be unstructured and leisurely. We may make participant observation or nonparticipant observation. Observation can be both overt and covert.

Evaluators want to make field observations to get a direct sense of the reality without an intermediary having to see and interpret it for us. Observation is not, however, a matter simply of opening our eyes and ears to people in real-life situations. We have to train our eyes and ears and must learn to record our observations. Diaries, check-lists, maps and diagrams, schedules, sociometrics, rating scales, and cameras can all be used for recording observations.

Observation schedules are by no means easy to write and a variety of errors are common in them:

Items	Comments
Is the student-teacher audible enough to pupils sitting at the back of the class?	Can this be "observed"? Or do we have to ask the backbenchers about it? Or should the evaluator walk to the back of the room and listen? The item needs clarification.
Does the student-teacher speak with confidence?	What should we look for when observing a display of confidence?
What economic status do the loanees have?	Can one "observe" economic status as such?
How did the loanees use the funds they obtained from the cooperative society: married second wives, paid children's school fees, engaged in heavy drinking, or bought new clothes?	How can we observe this history of behavior in a visit or during a short period of observation? Such information will have to be collected through alternative means.
An observation schedule seeks to observe: -- attitudes of people before the public meeting starts; and -- attitudes of the people during and after the public meeting.	Is this possible to do? Do attitudes change in the course of a public meeting? Do attitudes show on people's faces?
Does the cooperative society keep the books required under the law?	Okay, but isn't this a matter of an audit rather than observation?

Records and Documents

Records and documents are important sources of data for the evaluator. The analysis of records and documents may be quantitative (suited to the scientific paradigm) or qualitative (suited to the naturalistic paradigm).

The basic question asked from records and documents is "Who says what, to Whom, How, and with What effect and Why? The techniques of analysis are aggregation-integration method (quantitative), trend analysis (quantitative), content analysis (both quantitative and qualitative), and case-survey aggregation (qualitative).

We will have more to say about content analysis later in Chapter XIV.

The Ethics of Buying Data

The question has often been raised: Should an evaluator pay his or her respondents for participation in an evaluation study? There is no simple 'yes' or 'no' answer.

Knowledge production is a social function; and in the case of an evaluation study, the social use of evaluative information can often be quite clear both for evaluators and respondents. If the evaluator is working in behalf of the government or a non-profit making voluntary agency, it is public interest which is being served by the evaluation. The respondents, as good and concerned citizens, should freely participate in the evaluation study.

If, however, a subject is put in a position to choose between working on a construction site for the day or participating in your evaluation study, you should then pay to compensate for the wages lost by the respondent. But where the time invested could not have been sold for wages anyway, payments should not be made.

Summary

Evaluators typically use achievement tests, questionnaires, interview schedules, observation schedules, and records and documents as their sources of data. The evaluator using the scientific paradigm uses structured tools and instruments and formal approaches. The evaluator using the naturalistic paradigm makes himself or herself the instrument of data collection. The tools of the naturalistic evaluator are the unstructured in-depth interview, detailed observation and content analysis of documents and records.

The characteristic functions of the various evaluation tools of instruments have been described and problems in their design and administration have been listed.

Things to Do or Think About

1. Develop a detailed list of facts, principles, skills, and attitudes that you want your trainees to have learned by the end of your course.

2. Have you been interviewed recently by someone as part of an evaluation or a survey of some kind? What do you remember that was good about the interview? What did you find irritating or unacceptable? Was the interviewer able to win your trust?

3. Write an observation schedule on "Working Habits in the Office." Try it on a colleague. Ask your colleague to then try it on you.

CHAPTER XI

PRODUCT EVALUATION OR PRE-TESTING OF INSTRUCTIONAL AND TRAINING MATERIALS

Evaluation of instructional and training materials involves approaches and methods somewhat different from the evaluation of the effectiveness of development and training programs. A separate note on the subject of "product evaluation" is, therefore, in order.

Earlier, in Chapter I, we talked of two types of evaluation: formative evaluation and summative evaluation. The purpose of formative evaluation is to create evaluative information during the process of formulation of something, typically, a curriculum product - an instructional aid, a lesson plan, or even a total curriculum for a training course or a school grade. The purpose of summative evaluation is to sum up the effects of a curriculum or program or project and to discover what the summative impact has been and where. The problems of summative evaluation, since they typically relate to impact on learning behavior of individuals and communities have been well-covered in other parts of this monograph. In the following, we will focus on the formative evaluation of instructional products.

Product Evaluation

Product evaluation, also called pre-testing, has the objective of improving a lesson plan, a set of posters, a book for new literates, a film for pregnant mothers, or a social studies curriculum for a school district before it is finalized for release for more general uses.

The draft versions of a lesson plan, the typed manuscript of a book for new literates, the initial drawings of a set of posters, the rough-cut of a film are first used with a small but representative group of potential readers and viewers. This gives the teacher or trainer an opportunity to test his or her assumptions, the readability level of language used, learner perceptions of the graphics, the organization of ideas and other related communicational features of the material. The information thus developed is used in revisions, redesigns and rewrites of the material as follows:

Version$_1$------------Version$_2$-----------Version$_x$-----------Release for general use
Test$_1$ Test$_2$

A development trainer may be interested in the evaluation of any of the following instructional and training products:

Leaflets, folders, handbills
Literacy primers
Manuals, textbooks, followup reading materials
Illustrations for use with the above printed materials
Posters, charts, flannelgraphs, flashcards
Comics
Games and simulations
Audio programs of various kinds taped on cassettes
Puppet plays
Radio plays
Films, video cassettes and TV.

The Process of Product Evaluation

The process of product evaluation, or pre-testing, is about the same in case of all of the instructional and training products listed above:

1. The evaluator should produce a dummy, a replica, a rough print, or an approximation of the instructional material to be pre-tested which is as close to the final product as possible.
2. This rough correspondence to the final product should, again, be used with trainees in learning settings that approximate as closely to the real-life learning situations as possible.
3. A sample of learners should then be exposed to the instructional product and questioned about their perceptions, about what the product communicates to them, and whether or not they like using and learning from that particular instructional item.

While the steps involved in product evaluation or pre-testing may generally be the same for all instructional materials, the standards by which different instructional items will be judged will be quite different. (See the section on standards of quality for judging instructional materials in Chapter IX above.)

Product evaluation may sometimes involve more than one cycle of testing and revising. However, samples of learners used in pre-testing of instructional materials can be quite small.

Let us look at some examples of pre-testing below:

Example 1: Evaluating a lesson plan

Peter L. Higgs,[1] formerly of the Curriculum Development Center of the

---

[1] Peter L. Higgs, "How to Evaluate Individual Lessons," Lusaka, Zambia: Curriculum Development Center, Ministry of Education, Republic of Zambia, 1978.

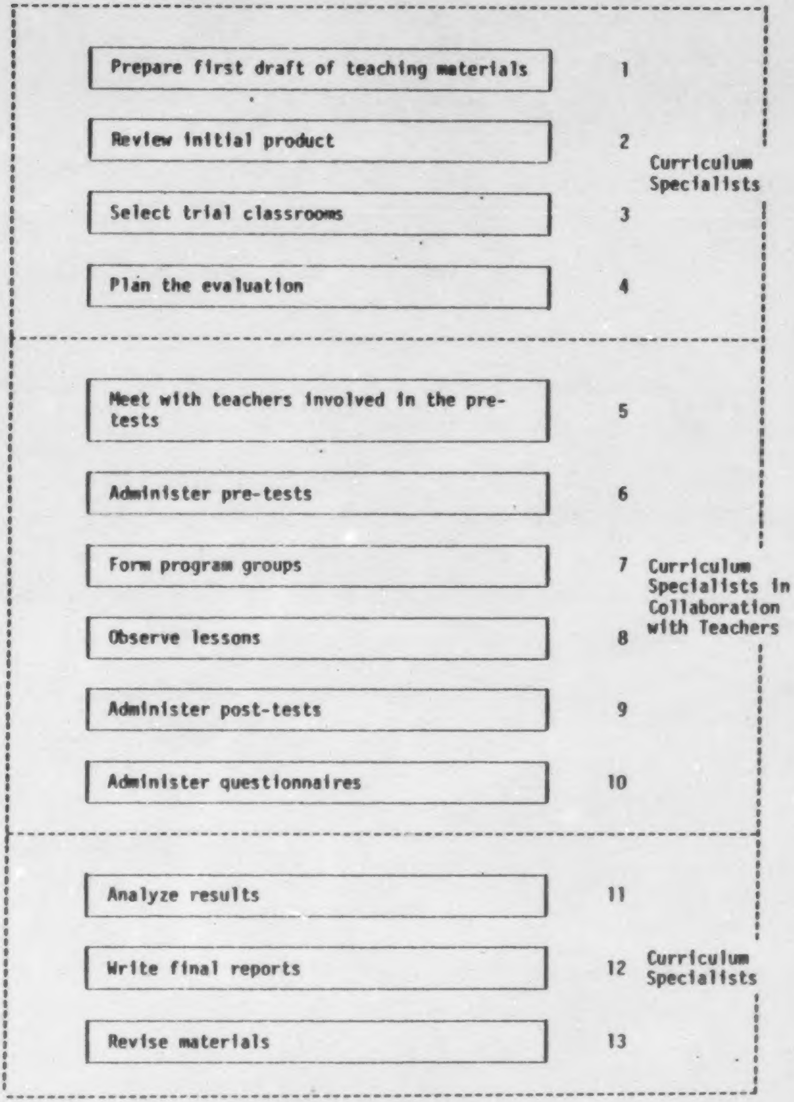
Government of Zambia and now of Unesco, suggested thirteen steps in the formative evaluation of a lesson plan. (See next page.)

Note step no. 6 that talks of administering pre-tests. Such pre-tests will, of course, have to be designed in terms of the teaching-learning objectives of the lesson. Typically, these will be written tests. These tests will help establish, what is called, the entry behavior level of a representative sample of learners.

Observation of the lesson (Step 8). Note that while the administration of post-tests is planned in a subsequent step 9, personal observation of the lesson is included in the pre-testing process. Personal observation is indeed a most important step that must not be missed. This observation may be formalized using observation schedules; or it may be somewhat informal and impressionistic. Whatever the approach used, without such observation, it will be impossible to know how the lesson actually "plays" in practice. How does the teacher handle the lesson? How do the students experience it? Are they able to relate the lesson to their everyday lives? What kinds of communication breakdowns seem to be occurring between the teacher and learners? What supplementation of the material might be necessary? Is the material, perhaps, too long for one lesson and must it, therefore, be broken into smaller units?

Use of the questionnaire in addition to a post-test. Note that in step 10, the application of a questionnaire is recommended. This is meant to generate that can be found neither through the post-test nor through observation of the classroom teaching. The evaluator must ask the teachers and the learners, as appropriate, about their inner experiences of working with the materials.

STEPS IN THE EVALUATION OF INDIVIDUAL LESSONS


Prepare first draft of teaching materials
1
Review initial product
2
Curriculum Specialists
Select trial classrooms
3
Plan the evaluation
4
Meet with teachers involved in the pre-tests
5
Administer pre-tests
6
Form program groups
7
Curriculum Specialists in Collaboration with Teachers
Observe lessons
8
Administer post-tests
9
Administer questionnaires
10
Analyze results
11
Write final reports
12
Curriculum Specialists
Revise materials
13

Report writing. The writing of the report recommended in step 12 is also important. A formal report not only creates a record, but also provides the evaluator with an opportunity to systematize his or her experiences; and draw deliberate conclusions and clear implications for the revision of teaching and training materials.

Example 2: Evaluating a set of posters

Some of the ideas discussed in the formative evaluation of a lesson plan under Example 1 above are directly applicable to the pre-testing of a set of posters. Some other ideas, to be discussed below, are unique to the evaluation of graphic materials.

Objectives. We will have to start with the question: What are the posters supposed to do? Are these posters teaching some information? Are these posters teaching particular attitudes? Are these posters teaching some skills? What essentially is the "message content" of the set of posters?

Clientele. A series of questions will also have to be raised in relation to the clientele of those posters. Whom will these posters teach the new information, the new attitudes, or the new skills? Are the clients literate or illiterate? Are they males or females or are they mixed groups? Are they members of particular clubs or participants of particular programs or are they members of the general public? This information will help us choose the pre-testing group.

Setting. Questions may also have to be raised in regard to the setting in which the posters will be seen. Will these posters be used in the classroom or will these be put on walls in the community center or in the church or even in the town market? If these will be used in the classroom setting, will these be integrated into lessons or will these be used as enrichment materials? The setting(s) in which the posters will be used must be recreated during the pre-tests.

Specimen materials. The posters to be evaluated will have to be prepared not in the final form, but as close to the final form as possible. This applies to size, graphic style, color, lettering, and even to the quality of paper on which the posters will be ultimately printed. Sometimes more than one versions of the set of posters may have to be prepared: with and without captions, line drawings versus photographs, etc.

Pre-testing. The first thing to do as part of the pre-testing will be to establish what the group of subjects chosen to look at the posters already know. Then the subjects should be given opportunities to view the posters under conditions very similar to those in which these posters will be later used or viewed. The evaluator should observe as the posters are being used; may have to give tests of knowledge to subjects; and may use questionnaires or interviews to supplement test score data.

What to test posters for. A set of posters can be tested in terms of at least three types of characteristics: (1) the integrity and the continuity of the set of posters as a set; (2) the graphic component of the posters; and (3) the verbal components of the posters.

The first question to ask in this case may be: Is the set of posters really a set? Do the posters go together? Do they add up to a message with a particular integrity? Do they have visual continuity?

As part of testing the graphic component, we will have to test the perceptions of viewers about the pictures. Do they understand the pictures? Do they find it possible to identify with them in a cultural sense? Do they like the style of graphics which may be photographs, line drawings, cartoon figures, stylized, or realistic? Do they like the color combinations? Questions about attensity (ability to attract attention) may also be tested here. What parts of a poster attract attention? What parts are missed or viewed only hurriedly? What parts are most remembered when the posters have been removed?

As part of testing the verbal component of the posters, we will have to check readability of the captions. Are the captions understood? Are they appealing? Is the lettering used of the right size? What do the captions add to the visual components?

Methods and samples. Is it possible to conduct fairly satisfactory formative evaluations of posters (and other instructional materials) using small samples of representative users. The writers and designers of such materials must, however, be themselves involved in these evaluations; they should themselves conduct the interviews and make personal observations.

Two methods have typically been used in pre-testing of instructional materials - the "tutorial method" and the "group-based method." In the tutorial method a single learner is involved at a time. Oral and written verbal behavior as well as non-verbal behavior of the single learner form the basis for revision and a rapid and subjective data analysis is undertaken. In the group-based method a representative sample of the target population is used in a group situation; a description of the aggregated group behavior forms the basis for revision and systematic statistical analyses of data are often undertaken. Both methods have been found to be equally effective. Indeed, no statistical differences have been found between the effectiveness of these two methods.

Summary

Evaluation of instructional materials (also called pre-testing of instructional materials, or simply product evaluation) presents a set of problems somewhat different from those faced in the evaluation of the impact of programs on the behavior of individuals and communities. The basic steps in the process of pre-testing of instructional materials are given and the need for a set of standards to judge quality of materials is indicated. Two examples are provided: one involving the pre-testing of a lesson plan, and another involving the pre-testing of a set of posters.

Things to Do or Think About

1. Adapt the "Steps in the Evaluation of Individual Lessons" given in this chapter to the evaluation of one of your own lessons.
2. Using the ideas included in this chapter (and particularly in Examples 1 and 2), can you develop a scheme for pre-testing an educational radio broadcast?

CHAPTER XII

DESIGNING EVALUATION STUDIES, AND THE CONCEPT OF DESIGN IN NATURALISTIC EVALUATION

In the dictionary meanings of the term, to design is to develop a conception of something, or is to prepare preliminary plans or sketches for something. In this sense of the word design, all evaluation studies must have a design. We must have a conception of what we want to do, why, and we must make some preliminary plans about how to go about doing what we want to do.

In the literature of research and evaluation, however, design has a highly technical meaning. To have an evaluation design is to follow a set of procedures that will increase the internal and external validity of evaluation results (when using the scientific paradigm); or the credibility and fittingness of results (when using the naturalistic paradigm).

In the following, we will discuss rather briefly the problems of evaluation design in both the scientific and the naturalistic paradigms.

Design in the Scientific Paradigm

Within the scientific paradigm, design typically means "experimental design". There has to be a sampling plan, and random samples must be obtained. Evaluation variables must be controlled through various mechanisms. Treatments should be well-defined and applied selectively to chosen samples. Instruments are often structured and statistical techniques are applied to the analysis of data collected.

It is beginning to be widely understood, however, that 'true' experimental designs are seldom possible in education and development. Random samples do not always make sense when dealing with special categories of subjects, in particular community contexts. Control of variables and treatments is often impossible. Evaluators are, therefore, now being offered "quasi-experimental designs" - evaluation designs that are half-way experimental. When using quasi-experimental designs, we try random assignment of treatments, if possible; but control when the data will be collected and from whom.

Design in the Naturalistic Paradigm

Within the naturalistic paradigm, they talk of aexperimental (non-experimental) designs. They do not seek to experiment upon individuals, groups and communities. The naturalistic paradigm demands that human behavior be studied as it occurs naturally, within its natural socio-cultural context. This means that neither the intervention of treatments, nor the control of variables and other contingencies is sought.

Thus, the design in the naturalistic paradigm has to be, what Guba and Lincoln have called, an emergent design or a rolling design. Plans and procedures for the evaluation study have to emerge within the realities of the field to meet needs as they are seen and to follow leads as they present themselves. Instrumentation has to be responsive to the data that happens to become available. Samples often get nominated and become exhausted when there is saturation of data and regularities in responses begin to show up.

Reliability and Validity

Researchers and evaluators working within the scientific paradigm swear by reliability and validity.

Reliability applies to a test or another measuring instrument. It is defined as a reasonable consistency in results obtained in a sequence or group of repeated tests and measures. A reliable test is one which gives consistent results in different applications to the same subject within a reasonable time-frame. Or, performs consistently when used by different evaluators, with different subjects. Reliability is necessary though not sufficient for validity.

Validity is the extent to which a test measures the thing it is supposed to measure. Support for validity may be logical or empirical. The test items may have been properly derived from accepted premises by rules of logic; or assumptions may have been based on supportable empirical evidence.

Internal and External Validity

The concept of validity applies not only to tests and instruments but also relates to the more general concerns of evaluation design. The results of an evaluation study and the conclusions drawn from these results must be seen as warranted, convincing and acceptable - that is, they must be seen as valid.

Let us look at some of the assertions that evaluators could make on the basis of their studies, and at the possible objections that could be raised to the validity of such assertions:

Assertions by evaluators	Objections to validity
The trainee group has shown considerable learning as evidenced by the high level of performance on the final test.	Maybe this group was familiar with the content of the training course even before joining the course. Maybe the test was easy or the grades have been inflated.

Adult attitudes towards literacy have changed drastically because of the project.	Maybe they have changed not because of the project, but because of the President's speech on national radio. Maybe they have changed not because of the project, but because of newly-opened textile factory has declared its preference for literate and semi-literate labor.
The group of farmers who undertook leadership training at the training institute had assumed actual leadership roles in the community more often than these farmers who did not join leadership training.	Maybe the farmers who undertook leadership training were already in leadership positions and wanted to increase their effectiveness as leaders. Maybe the farmers who joined leadership training were a self-selected group, fired with the ambition to capture the new leadership positions opening up in their communities. Maybe the other group of farmers that is not doing well, is different from the successful leadership group in important socioeconomic characteristics.
The farmers' training course increased the overall productivity of farmers who attended by 15% in a year.	Maybe the productivity increase for these farmers last year was 20%. Maybe similar farmer groups elsewhere have shown similar increases in productivity levels.
The introduction of the role of the Family Health Education Worker has changed the level of health in the selected communities from "Poor" to "Medium"	Maybe this is because of the heat and drought of the last year that killed all mosquitoes; and the famine relief high protein food aid that was provided to families in the area.

These are some examples of the assertions that could be made and the challenges to their validity. Professors Donald T. Campbell and Julian C. Stanley[1] have listed twelve different threats to the internal and external validity of evaluation studies. Evaluators should find their list most instructive:

---

[1] Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research. Chicago, Ill.: Rand McNally, 1963.

(A) Internal validity

1. History. An outside historical event, such as a presidential speech, or the enthusiasm generated by a newly announced economic plan could challenge the validity of the evaluator's claims.

2. Maturation. Individuals being tested as part of the evaluation may mature and grow in such significant ways that they may behave like different people by the time an evaluation study is completed.

3. Testing. The first test may teach the items on the test and other related and implied information. The same test (or an equivalent second test) may not then measure real changes brought about by the program.

4. Instrumentation. There may have been no changes in the reality but only in the calibration of instruments studying that reality. Or, different observers and examiners may have given different scores for the same unchanged reality.

5. Statistical regression. This is a statistical phenomenon. Extremely high or extremely low scores on a first test tend to move towards the mean of total scores during a second test. Thus, changes in the scores on a second test may really have nothing to do with respondent groups, program methods, or program effects. Statistical regression occurs specially in cases where groups have been selected on the basis of extreme scores.

6. Selection. Biases in the selection of learners for training, interviewing and testing may threaten the validity of results.

7. Experimental mortality. Those initially covered by an evaluation

study may cease to be participants in the evaluation. They may drop out of the program or may move away in search of food or work. Thus, the residual group may no more be representative of the group or community being studied.

8. Selection-maturation interaction. The peculiar chemistry of the selection process of subjects in an evaluation study and their maturation together may show effects independently of the program inputs and processes.

(B) External validity

9. The reactive and interactive effect of testing. The pre-test may increase or decrease the sensitivity or responsiveness of the respondent to certain program treatments applied as part of the evaluation.

10. Selection-treatment interactions. The peculiar chemistry of selection of respondents and the instructional and organizational treatments may create effects that falsify results regarding real program effects.

11. Reactive effects of experimental arrangements. Persons and groups show one set of effects of a treatment within the experimental setting, but not in non-experimental, real-life settings. Or, in some cases, experimental conditions may be much too artificial.

12. Multiple-treatment interference. When the same group is frequently tested, or interviewed many times in different connections, results may become confused. Effects of a test and an interview cannot be erased from the minds of respondents and the first test or interview may influence later testing and interviewing in ways that we do not understand.

The purpose of evaluation design is to reduce the above mentioned threats to the validity of evaluation results.

Credibility and Fittingness in Naturalistic Evaluation

In Chapter VIII (Models of Evaluation), we have reproduced a table developed by Guba and Lincoln comparing scientific and naturalistic terms appropriate to various aspects of rigor in research and evaluation. They pair them as follows:

Scientific term	Naturalistic term
Internal Validity	Credibility
External Validity/ Generalizability	Fittingness
Reliability	Auditability
Objectivity	Confirmability

Objectivity/Confirmability. Too often naturalistic inquiry is dismissed as *subjective* while scientific inquiry is considered *objective*. This is quite absurd because both types of inquiry (scientific and naturalistic) work with data provided by individual subjects. It is, therefore, always "subjective".

The "subjective", however, becomes "objective" when it is confirmed by a number of subjects or judges. Indeed, the significant word here is confirmability. If naturalistic data can be confirmed by others, then it is objective.

Reliability/Auditability. In naturalistic inquiry we can not talk of reliability because we are not dealing with structured instruments that are supposed to perform consistently. In naturalistic inquiry, the evaluator becomes the instrument. Instead of reliability, Guba and Lincoln have proposed the concept of auditability. Auditability means simply that the work of one evaluator can be tested for consistency by a second evaluator.

Internal Validity/Credibility. Internal validity gets translated into credibility in naturalistic inquiry. The evaluator can increase the probability of producing credible findings by allowing sufficient time for the novelty of his or her entry in the field to wear off so that respondents become familiar with the evaluator and the evaluator can check on own preconceptions; by establishing a proper rapport but by avoiding deep personal involvement; by checking on biases of both respondents and of self; by checking data for internal consistency through structural corroboration (each piece should validate the other and all pieces of data should add up to a plausible whole); and, finally, taking data back to respondents for their check.

External Validity (Generalizability)/Fittingness. The concept of generalizability is itself undergoing change and some researchers are questioning the very idea of context-free generalizations. In naturalistic evaluation the concept of generalizability (external validity) gets translated into the concept of fittingness. The evaluation findings from a naturalistic study should become the working hypotheses that would fit in similar programs in similar contexts elsewhere.

Some Ideas on Sampling

Validity and general rigorousness of evaluation studies can be increased by following proper sampling and design methods. We begin by presenting some simple ideas on sampling.

A sample is a portion, part or piece taken or shown as a representative of the whole. Sampling is often a practical need. Evaluators may deal with programs with broad scope, covering hundreds of thousands of people. They can not go to each and every member of their populations and ask them the questions for which they want answers. Instead they want to select a small number of respondents in such a manner that the sample is representative and can be studied to make inferences about the whole.

We should explain the two words population and representativeness used in the paragraph above. In everyday meaning of the term, population covers all the people - men, women and children, young and old, farmers, workers and housewives - living in a particular community or nation. For the evaluator, population is the total group of people in which the evaluator is interested. It may be all women in a country in the child-bearing age, all people suffering from lung diseases, all textile workers or all new literates in a region or a township. Samples are drawn from such populations.

Samples have to be representative, that is, as parts they have to represent the whole from which they are drawn.

There have been many advances in sampling theory. Statisticians have worked out formulas whereby they can test the representativeness of their samples and calculate the probabilities of error.

Size is an important consideration in selecting samples. Clearly the perfectly representative sample of a population is the population itself. Generally speaking, the larger the sample, the more representative it will be of the population. But unnecessarily large samples will not be good samples. We have to have the right size of sample that is both economical and representative.

On the next page, we have reproduced a table that can be used for determining sample sizes for various population sizes. Let us also look at some frequently used types of samples.

TABLE FOR DETERMINING SAMPLE SIZE FROM A GIVEN POPULATION[1]

N	S	N	S	N	S
10	10	220	140	1,200	291
15	14	230	144	1,300	297
20	19	240	148	1,400	302
25	24	250	152	1,500	306
30	28	260	155	1,600	310
35	32	270	159	1,700	313
40	36	280	162	1,800	317
45	40	290	165	1,900	320
50	44	300	169	2,000	322
55	48	320	175	2,200	327
60	52	340	181	2,400	331
65	56	360	186	2,600	335
70	59	380	191	2,800	338
75	63	400	196	3,000	341
80	66	420	201	3,500	346
85	70	440	205	4,000	351
90	73	460	210	4,500	354
95	76	480	214	5,000	357
100	80	500	217	6,000	361
110	86	550	226	7,000	364
120	92	600	234	8,000	367
130	97	650	242	9,000	368
140	103	700	248	10,000	370
150	108	750	254	15,000	375
160	113	800	260	20,000	377
170	118	850	265	30,000	379
180	123	900	269	40,000	380
190	127	950	274	50,000	381
200	132	1,000	278	75,000	382
210	136	1,100	285	100,000	384

Note: N is population size
S is sample size

---

[1] R. V. Krejcie and D. Morgan, "Determining Sample Size for Research Activities," Educational and Psychological Measurement, 30:607-610, 1970.

Random sampling. A random sample results when selections are made purely on the basis of chance, without any underlying system or pattern, and when each item or person in the population being studied has had an equal chance of being included in the sample. Random samples of appropriate size are most likely to represent all the characteristics and exact distribution of the total population of evaluator's interest. One method of taking random samples is to arrange the population in some way, assign numbers to it, and then draw some numbers randomly. Where the populations are big and the numbers to draw from are large, printed tables of random numbers can be used.

Random sampling may often be applied sequentially in evaluation studies. Geographical regions of a country may be selected randomly, followed sequentically first by the random selection of communities within the randomly selected regions, and then by the random selection of adults in the randomly selected communities. Again, randomly selected adults could be assigned to different learner groups through subsequent random selection.

List sampling. List sampling is a modification of the random selection method. The population of interest to the evaluator is arranged in a list according to some rule - alphabetically, for example - and then every nth number is selected from the list. For example, every 5th or every 20th number may be picked up depending upon the size of the population and the size of the sample being selected. The starting point in the selection process can be randomly selected to meet the criterion of equal chance of selection for each unit.

Area sampling. In area sampling, some geographical locations may be randomly selected from all available sites and then all appropriate units within the selected areas may be studied.

Stratified sampling. The population of interest to an evaluator may be divided into distinct socio-economic strata. Or, the population may be stratified according to age groups - children, young, middle-aged and very old. In such cases, stratified sampling may be used. In accordance with proportions in the total population, samples may be drawn proportionately and randomly from each of the population strata.

Purposive (theoretical or elite) sampling. The naturalistic evaluator or researcher may often need not a random sample but a purposive sample, a sample that fulfills his or her theoretically determined needs. The evaluator may be interested not in any randomly selected group of adults in a community, but in two or three people who are supposed to serve as the community's gate-keepers. The evaluator may be interested, that is, in small elite samples.

Some Simple Designs for Evaluators

A few designs of interest to evaluators are presented below. These descriptions are based on the work of Campbell and Stanley referred to earlier.

(i) The One-Shot Case Study

Campbell and Stanley call it a pre-experimental design. There is a total absence of control. A program treatment (X) is followed by observation (0):

X 0

While a case study implicitly compares its results with similar events casually observed or read and remembered, the case study can be strengthened by more systematic comparisons. At least one more comparison should be attempted. We should remember that this so-called pre-experimental design is the essential tool of the naturalistic evaluator.

(ii) The One-Group Pretest-Posttest Design

This is also considered a pre-experimental design and can be represented as follows:

$$0_1 \quad X \quad 0_2$$

A first observation or pretest (0_1) is followed by program treatment (X), after which a second observation or posttest (0_2) is recorded.

Evaluators will be using this design often in their evaluation studies. They should, however, do their best in defending their results against threats to their validity; or in qualifying their conclusions in the light of effects of history, maturation, testing or instrumentation as discussed above. (We have earlier discussed twelve threats to the internal and external validity of evaluation results. It will be a good idea for evaluators to develop the habit of checking their results in regard to each of these twelve threats, every time they design or complete an evaluation study.)

(iii) The Static-Group Comparison

It is a design in which a group which has been subjected to a program treatment is compared to another that has not been:

$$\begin{array}{cc} X & 0_1 \\ \hdashline & 0_2 \end{array}$$

This again is a design under many threats of validity. The most obvious ones are that of selection (the two groups may have been different to begin with) and mortality (subjects in the experimental group or the comparative group may have left the groups for some reason.)

(iv) The Pretest-Posttest Control Group Design

Campbell and Stanley call it a "true" experimental design. Two samples are randomly selected from the same population. One is assigned a program treatment and the other is not:

$$\begin{array}{lll} RO_1 & X & O_2 \\ RO_3 & & O_4 \end{array}$$

This design meets most of the standards of internal validity quite adequately, though care must be taken in generalization of results to the general population.

(v) The Posttest-Only Control Group Design

This is another example of the true experimental design. The pretest suggested in the design immediately preceding may not always be possible. It is not even necessary, if randomization in group selection can be assured. The design then takes the form:

$$\begin{array}{lll} R & X & O_1 \\ R & & O_2 \end{array}$$

(vi) The Quasi-Experimental Designs: The Time-Series Experiments

The time series design involves periodic measurement of some individual or group both before and after the introduction of some program treatment and the study of the "discontinuity" introduced in the pattern of behavior in time:

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$$

The evaluator using this design must specify in advance the expected time relationships between the introduction of a program treatment and the manifestation of its impact. The relative isolation of the group from outside influence should be ensured as well as some consistency in the conditions.

The above design can be strengthened by working with two groups in a time series as follows:

```
0    0    0    0     X    0    0    0    0
- - - - - - - - - - - - - - - - - - - - - -
0    0    0    0          0    0    0    0
```

(vii) The Nonequivalent Control Group Design

This is a design in widespread use because it fits the realities of the world of education and development as often faced. Too often evaluators have to work with already formed groups and classes and can not assign members to them randomly.

Thus, the design takes the form:

```
0     X     0
- - - - - - -
0           0
```

We should note the similarities between this quasi-experimental design and the "Pretest-Posttest Control Group Design" which was described above as a true experimental design. The essential difference between the two designs is that in the case of the "Pretest-Posttest Control Group Design" the treatment and the control group are chosen randomly while in the "Nonequivalent Control Group Design" discussed here, the groups are not randomly chosen and hence are nonequivalent.

Summary

The problems of reliability and validity (both internal and external) are discussed in relation to the scientific as well as the naturalistic paradigms. Twelve different challenges to internal and external validity as suggested by Campbell and Stanley are listed. There are no perfect designs that will take care of all the threats to the validity of evaluation results. However, appropriate sampling methods and choice of designs will enable trainers-evaluators to systematically consider rival explanations and to quality the results they do obtain.

Things to Do or Think About

1. Examine the conclusions of any evaluation study recently done by a colleague in your training institute or in some other development setting. What are some possible rival hypotheses or explanations for the assertions made by the evaluators?

2. Look at the table of "Assertions by evaluators - Objections to validity" included in the beginning of this chapter. What kinds of designs could have been used in each case to defend the validity of conclusions arrived at by educators?

CHAPTER XIII

IMPLEMENTING EVALUATION STUDIES: PROBLEMS OF DATA COLLECTION IN THE REAL WORLD

Many of the problems of implementing evaluation studies have been referred to directly or indirectly in other parts of this monograph. A systematic and self-contained discussion of the practical problems of conducting evaluations may, however, be more helpful and is included below.

Circumstances are sometimes stronger than men are. An evaluator can not control wind and weather, nor drought and famine. One can only cope with such circumstances and do the best possible. But many other possible sets of circumstances can be anticipated and one should be ready for them.

A New Set of Collegial Relationships

Evaluation is business unusual. Even when it is an evaluation of your own work by yourself, you perturb the existing relationships with your colleagues. It is important that you keep your feeling of self-importance in check and inform all concerned about what you are doing and why. Personal fears must be assuaged and professional jealousies must be relieved.

Evaluation will always make unusual demands on those who work with you in the office and in the field. The evaluator has to transform his officers, colleagues and assistants all into professional collaborators. The evaluator has to receive the blessings of those above; establish fair exchanges with those at the same level; and receive help from those below, not by ordering around but by sharing excitement as well as credit for the work done. Due acknowledgment must be made, both verbally and in writing, to those who provided advice or assistance.

Training of Field Investigators

In most cases, you as evaluator will not be able to collect all data single-handedly. You will need the assistance of colleagues and other field workers in the collection of data. It is important that those who have been mobilized as field investigators are provided appropriate training and orientation. The evaluator may not always want to inform the field investigators about the evaluation hypotheses or questions to keep out the personal biases of the field investigators. But the field investigators must be fully trained in the requirements of administering the evaluation instruments. Such orientation and training may have to be fairly extensive if in-depth interviewing is involved.

It is important that the evaluator is able to stay in constant touch with the field investigators to be able to answer their questions and solve unanticipated problems.

Piggybacking on Existing Institutional Resources

It is important that trainers-evaluators learn to piggyback on existing institutional resources. This is especially important in the case of transportation facilities. Travel arrangement to the field should be made to fit the travel plans of various officers from the parent department as well as other sister development departments.

Dealing with the Respondents

The evaluator can not anticipate famines and funerals, but must be aware of the seasons for migration of potential respondents, their daily patterns of work, and their festivals and hol day.

The investigator must be able to stay in the area long enough to wear off the novelty effect of his or her being there; to establish a rapport with the people; and to administer the questionnaires or to conduct the interviews.

The evaluator may have to use a third person to accompany him or her to conduct interviews with young mothers who may feel embarrassed being all alone with the investigator. In such cases, the third person will have to be chosen with care and the rules of conduct during the interviewing or questioning will have to be properly explained.

There will be situations when respondents will expect to be paid for being subjects of an evaluation study. As we have indicated elsewhere, evaluators (and researchers) should not pay for data unless a respondent will be losing wages in cash by participating in the evaluation study.

Changes in Samples and Instruments

In naturalistic evaluations, sampling is purposive and samples are developed and redefined to suit the circumstances. In the so-called scientific evaluation, samples are pre-determined and pre-selected. It will often happen that the evaluator is not able to collect data from the pre-selected sample and is obliged to make substitutions for the respondents lost or is forced to make-do with smaller samples.

It is not possible within the scope of this short chapter to deal with the complex issues of sample attrition and sample substitution. A general piece of advice can be offered, however. The advice is that evaluators must keep a precise and honest record of the changes made in the samples so that appropriate judgments can be made at the stage of interpreting data and results.

There will also be instances when changes in the evaluation instruments will be necessary. Some questions may not be understood by the respondents in an evaluation study. Some questions may be unanswerable, and some others the respondents may refuse to answer. The evaluator should be in touch with the field investigators (where field investigators are involved) to discuss problems and make the necessary changes. Changes made in the instruments should be followed uniformally by all field investigators. Clearly, such situations can be avoided by proper pre-testing of evaluation instruments.

Handling of Completed Instruments

Problems can arise from careless handling of completed instruments. Questionnaires and interview schedules can get lost. Data is precious and should be treated as such. Field investigators should be instructed clearly in regard to mailing and despatch of data. Should it be sent by hand with officials travelling from the field to the city office? Should it always be mailed? How should it be packed? Should it be sent by registered mail?

Completed questionnaires and instruments can get mixed up in the evaluator's office. These should be properly marked and coded as soon as received.

Summary

Various problems can arise in the course of actual implementation of an evaluation study. Colleagues may not collaborate. Field investigators may mishandle the administration of evaluation instruments. Getting to the field may be difficult. Respondents may be inaccessible and instruments may need last-minute changes. Finally, collected data may be lost through careless handling. While an evaluation study may never be completely problem-free, the evaluator should be able to anticipate some of these problems and fix others as they appear.

Things to Do or Think About

1. What are some of the problems that you anticipate in the course of data collection in your setting?
2. What are your suggestions for evaluators in regard to establishing fruitful collaborative relationships with their colleagues and subordinates?
3. Can you think of cases where problems in data collection in the field killed an evaluation study?

CHAPTER XIV

THE PROCESS AND TECHNIQUES OF DATA ANALYSIS

After tests and questionnaires have been administered, interviews and field observations have been conducted, and relevant records and documents have been collected, what we then have is raw data. Raw data, in itself, is not information. Information has to be systematically produced from available raw data for use in programming decisions.

The Meaning of Data Analysis

The process of converting the raw data into evaluative information is called data analysis. The skills and techniques of data analysis involve coding and scoring, combining and weighting, standardizing, ranking, tabulating, making graphs, developing crossbreaks and running statistical tests for correlations and differences. Or, the process of data analysis may require developing case studies and profiles; thematic analyses and writing analytical reviews or product evaluations of curricular and training materials. We will deal briefly with most of these approaches and techniques in this chapter. Statistical analysis of data will be discussed separately in Chapter XV.

Some Tools of Data Analysis

The following are some of the tools of data analysis:

1. A good supply of ruled and plain paper
2. A supply of lead pencils and a pencil sharpner
3. Erasers

4. If possible, a bottle of liquid paper (or another correction fluid)
5. A pair of scissors
6. Scotch tape with dispenser
7. Paper clips and pins, and
8. A set of colored pencils.

A Basic Suggestion

In the process of data collation and analysis, write only on one side of the paper. Use a separate sheet of paper for each single idea and theme that you develop. This will help you later in trying different organizations of the material. You do not have to use nice and fresh paper for this stage of data processing. You should use discards from cyclostyled materials and any other scrap paper you can get hold of. For making tables by hand, use ruled paper so that rows of data can be read without confusion. Be careful about the spacing of numbers in columns:

```
125                 125
  5                 5
 11 is correct      11  is not correct
```

Do not write over your own writing. Use an eraser; or strike out and write afresh.

Clustering and Identification of Data Pieces

For the sake of convenience, let us call the tests, interview schedules, observation schedules and questionnaires filled and returned by investigators and respondents, <u>data pieces</u>. The very first thing to do when all the data pieces are in, will be to arrange and identify the various pieces by numbering them. Different arrangements will be appropriate in different cases. Where respondents are not anonymous, pieces may be arranged alphabetically. Other arrangement may reflect clusters of data pieces; and these data pieces may be clustered by sex, age, religion or ethnicity; by training course, batch or year; by region, province or district; by literacy teacher in charge; in terms of trained versus untrained groups; and by training methodology used.

An example of data pieces organized by clusters may look like the following:

Region X

- Primer I
  - 1 Male 1
  - 2 Male 2
  - 3 Male 3
  - 4 Female 1
  - 5 Female 2

- Primer II
  - 6 Male 1
  - 7 Male 2
  - 8 Female 1
  - 9 Female 2
  - 10 Female 3

Region Y

- Primer I
  - 11 Male 1
  - 12 Male 2
  - 13 Male 3
  - 14 Female 1
  - 15 Female 2

- Primer II
  - 16 Male 1
  - 17 Male 2
  - 18 Female 3
  - 19 Female 4
  - 20 Female 5

Such clustered organization and identification of data pieces helps at the later stages of data analysis.

Once organized according to need, all data pieces should be given permanent numbers in the upper right-hand corner on the face of each piece, using a color code.

If a whole set of instruments -- an achievement test, an interview schedule, an observation schedule and a questionnaire -- have all been used with the same one group of respondents, then a matching numbering system should be used. For example:

Name	Interview Score (I)	Observation Score (O)	Questionnaire Score (Q)	Test Score (T)
Abram	I-1	O-1	Q-1	T-1
Binti	I-2	O-2	Q-2	T-2
Camaro	I-3	O-3	Q-3	T-3
Daudi	I-4	O-4	---	T-4
Elice	---	O-5	Q-5	T-5
Fakouri	I-6	O-6	Q-6	T-6

Make sure that you write I and T and O and Q clearly enough so that one is not confused with the other. Note that in the above display, Daudi's questionnaire is missing, as is Elice's interview schedule. However, Fakouri still gets numbers I-6, O-6, Q-6 and T-6 for his data pieces. In other words, all data pieces for the same one person get matching numbers.

The Need for Immersion in the Data

After the data pieces have been arranged and numbered, it will be time to do two further things: to recollect the evaluation questions that needed to be answered by the evaluation study; and to get immersed in the data already collected.

Write out the list of questions you want the data to answer. If there are sub-questions to the questions, write them out also. For example, an evaluator may write his or her questions as follows:

1. How are trained assistant adult education officers different in regard to their overall performance from untrained assistant adult education officers?

    1.1 How do they differ in regard to their technical knowledge about development and adult education?

    1.2 How do they differ in regard to their knowledge of the literacy methodology being used in the program?

    1.3 How do they differ in regard to their supervision styles, and diagnostic and problem-solving skills?

    1.4 How do they differ in terms of their attitudinal orientation to adult learners, rural communities and their own work?

Armed with such a list of questions, it is time to begin the immersion in the data. By immersion in the data, we mean going carefully through all the pieces of data, piece by piece, page by page, item by item; studying all the responses; and making careful, written notes. You should take note of the expected, of the unexpected and of the curious; of the emergent pattern and of the seeming relationship, as you go through the data. This immersion may require more than one dip; that is, you may have to do more than one reading of the data pieces. The time used in going through this process is always well-spent. Therefore, be patient.

Possibilities and limitations of the data collected. This will also be the time to discover the unanticipated possibilities of the data. For example, a questionnaire used with students of agricultural extension to evaluate their attachment experience, may be full of information about prevalent practices on butchering meat animals; or on the popularity of poultry farming in a particular region. On the other hand, serious problems may be discovered with the data during the immersion process. Some questions in the test may have been consistently misunderstood. Some other questions may have received "socially acceptable" responses, not the real answers.

Some pieces may have to be discarded altogether for being incomplete or dishonest. It may become clear to the evaluator that available data will not make an overwhelming case for or against a particular position or approach; and the evaluator may have to warn readers against drawing unwarranted conclusions. All this should be taken note of during the process of immersion in the data.

Developing Models of Data Analysis

In the process of data analysis, we have to work back and forth from both ends -- from the direction of the raw data we have and from the side of the final information we need to produce.

To make sure that we do not get lost in the details of the data collected, we must keep in mind the questions that will ultimately have to be answered. We must also develop a model or plan of data analysis. This so-called model of data analysis, may be nothing more than a reincarnation (rebirth) of the evaluation design used earlier in conducting the evaluation study.

In Chapter XII, we had listed some of the typical designs used by evaluators: simple comparisons between groups, before and after designs, etc. Some evaluation studies may not fit the experimental or quasi-experimental modes discussed in Chapter XII. Some evaluators, for instance, may be interested in the study of patterns of perceptions and configurations of relationships in the field. Given below are some examples of models of data analysis actually proposed for use by trainees in DSE workshops on evaluation of basic education and development training programs:

Example XIV.1

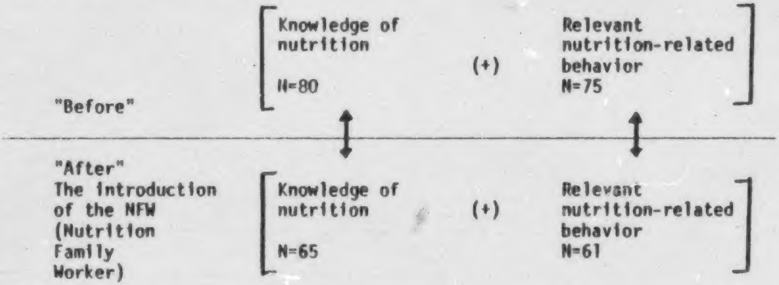
The model of analysis involving comparisons between two groups across a set of characteristics may be quite simply presented as follows:

	Characteristics	I	II	III	IV
Group A N=60					
Group B N=51					

Once data has been collated and put in order as shown above, the process of analysis will become easy to undertake. Some similarities and differences may be clear even through simple visual inspection. Some differences may have to be tested statistically. (See Chapter XV below.)

Example XIV.2

A "before and after" evaluation design may now appear as a model of data analysis as follows:

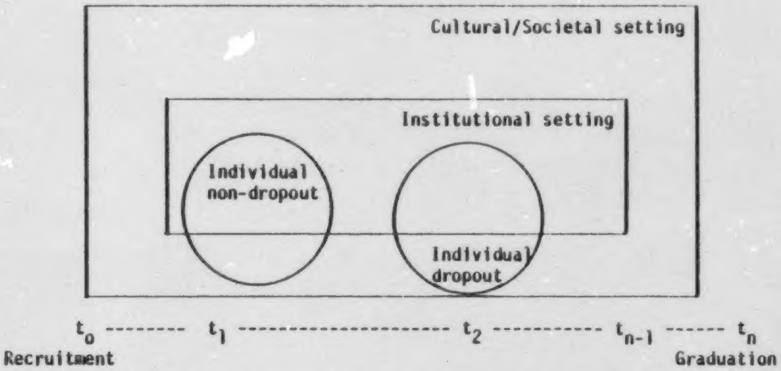

"Before"
Knowledge of nutrition
N=80
(+)
Relevant nutrition-related behavior
N=75
"After"
The introduction of the NFW (Nutrition Family Worker)
Knowledge of nutrition
N=65
(+)
Relevant nutrition-related behavior
N=61

Note that the sample sizes can vary and can be shown in the model for the reader's information.

Once again, projecting a model (or plan) of data analysis as shown above is helpful in the processing of data and its adequate analysis to get the needed answers.

Example XIV.3

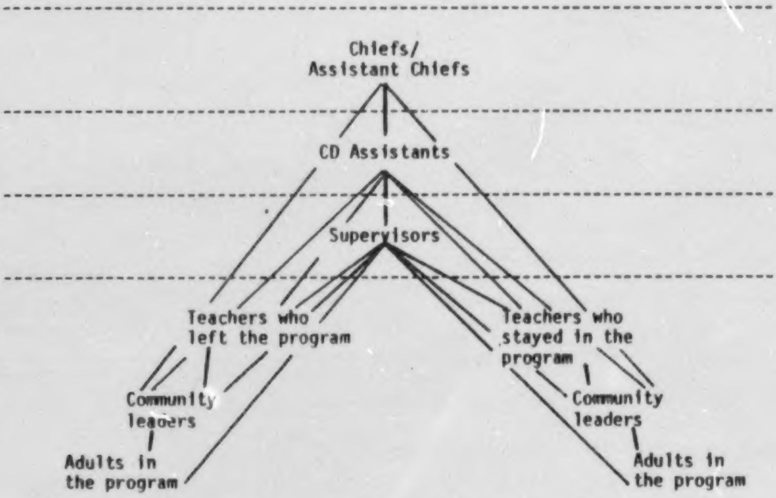
In some evaluation studies, the group A versus group B, or "before and after" analytical designs may not be applicable. System mapping (or what I have elsewhere called configuration mapping or plotting) may be needed for collating data, for understanding patterns of relationships and for the analysis of data. An evaluation study may ask the question: Why do students drop out of a university training course and may conceptualize analysis and interpretation of data in terms of the following configurations:


Cultural/Societal setting
Institutional setting
Individual non-dropout
Individual dropout
t_0 -------- t_1 -------------------- t_2 ----------- t_{n-1} ------ t_n
Recruitment
Graduation

The non-dropout and the dropout may be compared with each other in terms of individual characteristics, and in how they might respond differently to institutional and social contexts. The phenomenon of dropping out could also be made time-sensitive to identify any phases and cycles inherent in the process. Without developing such a model of data analysis, the evaluator may not place individual data in context, may fail to ask some exciting questions and may not obtain some most useful answers.

Example XIV.4

Finally, a configurational map -- a combination of the organizational chart and a sociometric network -- may be used as a model of data analysis in cases where actors at different levels of an enterprise are asked the same or similar questions on the various aspects of a program. In an exploratory study dealing with the turnover of literacy teachers in a literacy program, an evaluator may ask a similar or the same set of questions from a wide range of people, such as, chiefs and assistant chiefs, community development (CD) assistants, supervisors, community leaders, teachers still teaching in the program and the teachers who dropped out, and adult learners in the program. The model for data analysis may be drawn as follows:


Chiefs/
Assistant Chiefs
CD Assistants
Supervisors
Teachers who left the program
Teachers who stayed in the program
Community leaders
Community leaders
Adults in the program
Adults in the program

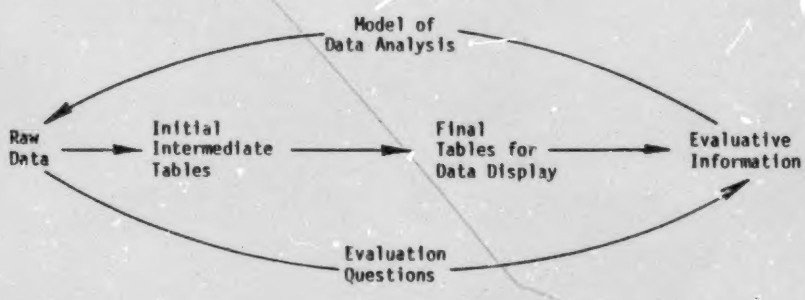
Such a configurational map may help both in the organization of data (what to say first, what to say next, etc.) and in the analysis of data as we move across levels and deal with different constituencies in the social network, with their special perceptions and interests.

Before taking leave of this topic, we should indicate that the "set of evaluation questions" and the "model of data analysis" are two different versions of the same organizing principle. Each should clarify and complement the other. Both may be used by the evaluator in the process of data analysis.

Dummies of Intermediate and Final Tables

As we have suggested earlier, one needs to work back and forth between raw data and the needed information as part of the process of data analysis. This going back and forth will involve, as we have suggested above, a recollection of the evaluation questions and the development of a model (or plan) of data analysis. It will also involve the design of dummies (or samples) or tables. Some of these tables will be used for the display of data in the final report; but others will be initial and intermediate tables which will be used in the processing of raw data but will not appear in any report.

The whole process of data analysis discussed so far can be visualized as follows:


Model of Data Analysis
Raw Data
Initial Intermediate Tables
Final Tables for Data Display
Evaluative Information
Evaluation Questions

Processing Raw Data:

Coding, Scoring and Standardizing

The first important step in the process of data analysis may often be coding and scoring.

Coding

Coding means to assign a particular code to a particular category of responses. The following are examples of codes:

Code 1	Prefers condoms as family planning aids	A
	Prefers an IUD for his wife	B
	Prefers to do family planning by abstinence	C
Code 2	Has insufficient (low) nutrition information	L
	Has average (medium) nutrition information	M
	Has high degree of nutrition information	H

Scoring

Scoring is assigning numerical values to particular responses or to particular levels of performance. Attitudinal responses will be often qualitative and will need to be scored. The same is true of performance scores which may involve observation of performance, judgment on what is observed, and, ultimately, the change of judgments into some sort of quantitative scores.

Standardizing

To standardize scores is to so treat them that they can be compared using the same one yardstick. A profit of 75 shillings on a 400 shilling investment is not easily comparable with a profit of 15 shillings on a 50 shilling investment. When both profits are standardized as percentages (18.75% versus 30%) they are easily comparable.

Evaluators of development training programs will often have to compare scores made by individual trainees on a variety of achievement and performance tests. Each time scores are to be compared, the evaluator should check if prior standardization of the scores will be necessary.

In the examples that follow, the processes of coding, scoring and standardization have been demonstrated:

Example XIV.5

In evaluating the effectiveness of a training program for agricultural teachers, a classroom observation schedule used the following items:

TEACHING SKILLS

(i)	Provides introduction to the lesson	Right	Wrong	Confusing
(ii)	Changes method according to need	Yes	No	Reluctantly
(iii)	Helps students recapitulate the lesson	Periodically	Not at all	Only at the end
(iv)	Accepts and answers questions	Always	Not at all	Sometimes
(v)	Gives individual attention	To all	To none	To some poor students
(vi)	Helps the students writes notes	Always	Not at all	Sometimes

(Not all of the items under the heading "Methods," in the original schedule, have been used above.)

We do not wish to make any comments here on the merits or demerits of the items as written. The point we want to make here is simply that some numerical values must be assigned to the judgments made during the observation; and that those values must be aggregated for use in data analysis. For example, approved behavior may be assigned a score of +1, an indifferent behavior may be assigned the value of 0, while an unacceptable behavior (which will hinder learning) may be assigned a value of -1. This will enable the evaluator to come up with an aggregated score for the teaching skills, as suggested below:

(i)	+1
(ii)	-1
(iii)	+1
(iv)	0
(v)	0
(vi)	+1
Total score:	2

It is important to note that different types of questions can be asked from the same data. For example consider the question: Do student-teachers, typically, help children to recapitulate ideas given in a lesson? Looking at answers only to (iii) above, for all the student-teachers tested, an answer to this question can be found.

Example XIV.6

A questionnaire (filled by each student individually, but sitting as a group in a large hall) sought to evaluate the effectiveness of field attachment of agricultural students. Questions were asked on a variety of themes, such as:

(a) Disease control

(b) Meat hygiene

(c) Dip management and tick control

(d) Artificial insemination service

(e) Veterinary investigation

(f) Clinical centers

(g) Farm visits

(h) Extension and education services

(i) New skills learned

(j) Show and field days

(k) Management practices

(l) Unique farming problems

(m) Individual problems, and

(n) Field attachment experience.

Each of these themes, from (a) to (n), had included a set of questions. For example, the theme (a), Disease control, had been broken down into the following items:

1. What notifiable disease(s) did you come across?
2. What methods of control and prevention were used?
3. Mention the vaccination campaigns you saw.
4. Enumerate the diseases against which vaccination was done.
5. What were the reasons for vaccination?
6. How was the vaccination organized and carried out?
7. How was the vaccine administered? Indicate any special precautions taken.

8. How many animals were vaccinated?
9. What was the dosage of the vaccine?
10. What was the cost of the vaccine per dose?
11. How did the farmer pay for it?
12. What is the duration of immunity for the vaccines used?
13. What was the type of vaccine used -- live, attenuated or dead?
14. What were the problems encountered in the vaccination campaign?
15. How were the vaccines used handled?

As we can see, these questions are a combination of (i) knowledge of the student of technical information; (ii) recall of "what" was done and "why" in some problematic situation in the field; (iii) information about some local happenings during the period of the student's attachment; and (iv) descriptions of professional actions and technical practices seen by the student during the field attachment but on which the student might have had no control.

In this case, the evaluator will first have to separate items of student's responsibility from those items which were part of the context; and then will have to make judgments about the quality of student performance in the given circumstances. The evaluator may assign A,B,C,D and F grades (or some number grades) to the performance of each student.

Once, again, we should note that many different uses can be made of this data, in addition to evaluating student performance. Using the same data, one could develop evaluations of dip management or clinical centers in the country; learn about the diffusion of new skills within rural communities; or learn about farm management practices, in general.

Example XIV.7

In an evaluation of the on-the-job performance of cooperatives storemen, the interview schedule to be used with the storemen's supervisors, had the following item:

		Excellent	Good	Fairly Good	Poor	Very Good
Knowledge and application of Code of Regulations	BEFORE				X	
	AFTER			X		

Notice that data produced by this item will have to be processed into an "improvement" score with a numerical value, before it can be further analyzed to study correlations and patterns of inputs and consequences in a training setting.

The problems of scoring interview and observation data, to change qualitative into some kind of quantitative data, can not be completely eliminated in this value-laden world of ours. However, some serious problems can be mitigated at the stage of instrument design and item construction. Tools, and items included in those tools, can be so designed as to elicit answers that are more easily amenable to quantification.

Weighting and Combining Scores

As teachers we know that in writing achievement tests we can assign different marks to questions on the question paper, depending upon the difficulty or the importance of particular questions. This differential allocation of marks to different questions (and answers) is called weighting.

Weighting is also involved in the analysis of opinion and attitude questionnaires and observation schedules. Needless to say that allocation of weights to responses on an attitudinal scale should be undertaken with care, especially in regard to the values of neutral, positive and negative responses.

The "Blankets" and Crossbreaks

After the coding, scoring, standardizing and weighting have been completed, it is time for the evaluator to have a full and complete, overall look at the data. This can be done by developing large-size tabulations that show at one glance the responses made by all the different subjects on a total test, a whole questionnaire or some other instrument.

Subjects / Test Items	1	2	3	4	5	. . .	20	. . .	30	. . .	n-1	. . .	n
1													
2													
3													
4													
5													
⋮													
25													
⋮													
50													
⋮													
100													
⋮													
N-1													
N													

Such tables could include a hundred or more subjects and their responses on 30, 40 or 50 items on a test, a questionnaire or an interview schedule. These could be big tables and cover half the surface of a wall. The participants of the DSE workshops in Kenya had started calling these tabulations "Blankets," a name we have retained.

Trainers-evaluators will most often be able to accommodate their data on one or more blankets -- one each for a test, a questionnaire and an observation schedule, for instance. Where the sample size is in many hundreds or even in the thousands, the evaluators will have to use more sophisticated methods, including the help of a computer for data analysis.

A careful look at a blanket would suggest many different leads to the evaluator in regard to response patterns, and differences and correlations between items. By consolidating data in blankets and by focussing on the various rows and columns of the larger blanket, one can develop many useful crossbreaks.

Crossbreaks are smaller blankets and can serve useful functions. Examine the two crossbreaks reproduced below:[1]

---

[1] H. S. Bhola, Evaluating Functional Literacy. Amersham, Bucks, U.K.: Hulton Educational Publications, 1979.

Lesson	I					II					III				
Question Number	1	2	3	4	total	1	2	3	4	total	1	2	3	4	total
Total possible points	5	5	5	5	20	4	5	6	5	20	6	5	3	6	20
Learner A	5	3	2	1	11	3	2	2	1	8	4	4	3	5	16
B	5	3	4	4	16	4	0	5	4	13	5	5	3	4	17
C	4	2	5	4	15	4	4	5	5	18	6	4	3	5	18
D	5	3	1	2	11	2	3	2	1	8	4	4	3	5	16

FIGURE 1. *Data on a mathematics test*

Lesson	I			II			III		
Attendance	High	Medium	Low	High	Medium	Low	High	Medium	Low
Learner A	11				8				16
B		14				13		17	
C			15			18			18
D	11			7					16

FIGURE 2. *Attendance and final grades crossbreak*

In the first crossbreak above, see under Lesson 1, items 1 and 2. All learners have done almost equally well on item 1, and equally poorly on item 2. Under items 3 and 4 under Lesson I no patterns seem to emerge. Maybe items 1 and 2 are not good items since they do not help us separate good students from bad ones. Or, maybe items 3 and 4 are poorly written and need to be reworked.

In the second crossbreak, the changes in performance across lessons I, II and III can be clearly seen, and different interpretative questions can be raised through simple visual inspection.

Data Display by Tabulation

Tables to be useful must be accurately compiled and should be easy to read and interpret.

In the local library or in a bookstore, you may be able to find some manual for writers of term papers, theses and dissertations. These manuals provide excellent help on how to compose tables. It is not within the scope of this document to provide detailed instructions on how to make tables. We will be satisfied with making the following general suggestions:

1. Number your tables as TABLE,1; TABLE,2, etc.
2. Give a title to each table; and make the title both accurate and complete.
3. The headings and descriptions used for rows and columns should also be accurate and complete.
4. Use correct placing and spacing, especially where numbers and decimals are involved.
5. Do not make up your own abbreviations. Use only standard abbreviations. Even when standard abbreviations are used in a table, explain them in the footnotes to the table.
6. Sometimes, statistics from different years may have to be used in the same one table. Indicate which year those statistics belong to, e.g.:

Population figures in millions (1975)	Per capita income (1974)	Radio sets in use (1975)

7. Separate "estimates" from "actual counts." Do not confuse, one with the other.

8. Wherever necessary, qualify your data. For example, you may have to say: Figures do not include data from Korea; or Domestic workers have not been included, etc.

9. Standardize your scores, if at all possible. However, if standardized scores are misleading, also include absolute scores.

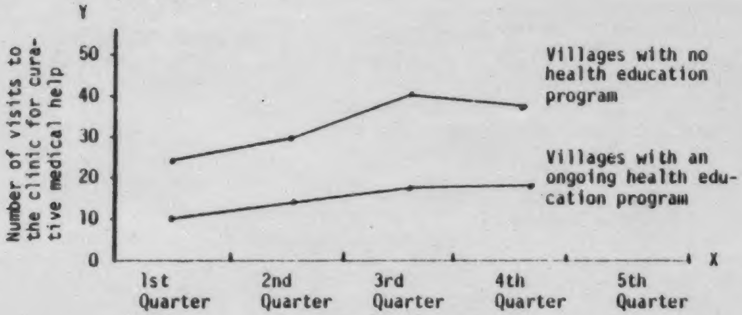
10. Sometimes, comparative statistics may have to be included in tables to make sense out of a given set of statistics. One can get a better idea of the level of poverty in a country by seeing, in the same table, the per capita income figures from U.S.A. or Sweden or even from a richer neighboring country.

Data Displays by Graphics

The question of preparing graphics for displays of data is important. Graphics communicate ideas simply and attractively, but they are not always easy to make. There is a lot to learn about making graphics. It may interest the readers that there is a special national Council on Social Graphics in the Bureau of Social Science Research in Washington, D.C. which recently held a general conference on the topic of "Graphics for Data Analysis and Social Reporting." It is not within the scope of the present paper to discuss the preparation of graphics for data display at any great length. Some brief comments are included here on the three most basic graphic displays of possible use by trainers-evaluators: (1) the line graph, (2) the bar graph and (3) the pie chart.

The line graph

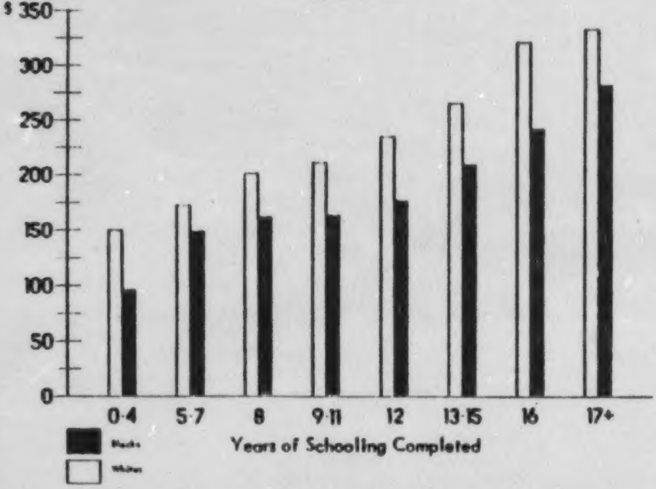
Examine the line graph included below. The O X axis has been divided into a time scale of .8 centimeters for each month and, therefore 2.4 centimeters



for a quarter. The O Y axis is divided into parts by .8 centimeters, each part representing ten visits to the clinic. These scales are chosen according to convenience. It is helpful to use graph paper for the initial construction of line graphs (as well as of bar graphs). The graphs can later be copied or traced to suit various purposes. As can be seen, line graphs can be used to present data somewhat dramatically to make a strong impact.

The bar graph

Bar graphs are easy to construct and can make effective displays of data. Examine the bar graph below. The data about years of schooling as it interacts with racial origin in the U.S. is presented quite effectively.



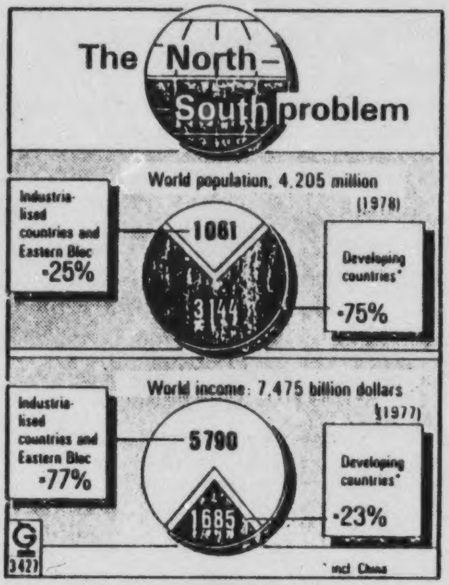
Blacks and Whites: Weekly Earnings, 1970
(Males, aged 35–44)

Chart by World Education, 1978 (Based on data from NACAE, 1974, p. 33).

The pie chart

Finally, some data may be presented effectively through a pie chart. Examine the pie chart below.

BEST COPY AVAILABLE



Reproduced from Development and Cooperation, Journal of the German Foundation for International Development, No. 5, 1981 (October/November), page 7.

The pie charts are a little more complicated to draw than bar graphs. The reason is that we need to work with proportions of 360, the total degrees of the angles projecting from the center of a circle. For example, in the upper pie chart:

$$1061 : 3144 \quad \text{is}$$

$$\frac{1061 \times 360}{(1061 + 3144)} : \frac{3144 \times 360}{(1061 + 3144)} \quad \text{or}$$

$$91 : 269$$

Similarly, in the lower pie chart

BEST COPY AVAILABLE

$$1685 : 5790 \quad \text{is}$$

$$\frac{1685 \times 360}{(1685 + 5790)} : \frac{5790 \times 360}{(1685 + 5790)} \quad \text{or}$$

$$81 \qquad 279$$

To make pie charts one would also need to have a compass to draw circles and a protractor to mark out angles and sectors of the circle.

Problems of Data Analysis in Naturalistic Evaluation

The problems of data analysis in the context of naturalistic evaluation are quite unique. Here, the evaluator would have collected detailed notes (rather than numbers or responses that can be quantified) on the basis of in-depth interviews and unstructured observations, and must make sense of all that qualitative data. This will involve content analysis, thematic analysis and process reconstruction.

Content analysis. According to Holsti content analysis is "any technique for making inferences by objectively and systematically identifying specified characteristics of messages." The aim is to categorize. The categories should be exclusive, exhaustive, internally homogeneous, form an integrated set, be reproducible, credible and must answer the questions the evaluator has to deal with.

Thematic analysis. As the name suggests, the evaluator must read transcripts of the depth interviews and observations for the themes that pervade the data and then use the data to test the hunches and hypotheses suggested by the themes.

Process reconstruction. Where change is the focus of an evaluation, the evaluator must try to reconstruct the process by which change could have occurred and the dynamics that might have been involved. This would involve systematic "cause hunting."

In all of the above cases, there will be the need to be theoretical. Some theory, model or conceptualization must inform our data analysis. In the case of overall change, for example, a change model[1] could be used to develop before and after profiles of field realities and to identify the intervention that lead to change.

Discussion of Results

In the following, some general suggestions are made about discussion of results obtained from data analysis:

1. Relating with preconditions and entry behaviors. As part of the discussion of results, reexamine available data on entry behaviors and study the preconditions that prevailed when the training episode of your interest began. The phenomenon of high dropout rates from a college course, for example, may be explained better in terms of faulty recruitment methods than by what is taught during training. The failure of a family planning program may be explained better in terms of the precondition of extremely high infant mortality rates in the region.

2. Putting things in context. Analyze findings in terms of the institutional and the social contexts of training programs. Do some institutional policies actually go against policies of rehabilitation of distressed families or against increasing individual savings? Does the social context promote or inhibit cooperative behavior?

---

[1] H. S. Bhola, "Planning Change in Education and Development: The CLER Model in the Context of a Mega Model," Viewpoints in Teaching and Learning, Vol. 58, No. 4, 1982.

3. Relating with whât is already known. Compare and contrast what your data tells you with what is already known. Do your findings surprise you? Are your findings reinforced by what other evaluators have found in other settings? What was expected? What is unexpected?

4. Looking for correlations and causations. Data analysis will typically involve search for correlations and even for causal links. In so doing, think of the rival hypothesis -- an alternative explanation for what you see. Consider all possibilities before making broad assertions.

5. Reexamine your assumptions. It is important to keep on thinking about the assumptions on the basis of which the evaluation study was designed and the evaluation questions were raised. Did those assumptions hold up? How have those assumptions changed.

6. Relating to the limitations of data. Discuss results in terms of the limitations of data discovered, as the evaluation design was implemented and evaluation tools and instruments were actually used. Some limitations of data may indeed be fatal to the study and to the conclusions drawn from it. Another set of limitations may be less severe, but may introduce the need for a high degree of caution in interpreting results of an evaluation study.

7. Setting up norms for success and failure. The evaluator must establish norms for success or failure of a program being evaluated. What kinds of results will provide the cause for satisfaction? What results will be interpreted as failure?

Summary

In data analysis one needs to go back and forth between the raw data obtained and the evaluative information that needs to be produced. This will require coding, scoring, standardizing and weighting of raw data. It will also require preparation of appropriate tables and graphics for the display of data. Naturalistic qualitative data requires a kind of analysis different from quantitative data. In both cases, it will be helpful to have a model (or plan) for data analysis to assist in the process of producing useful evaluative information.

Things to Do or Think About

1. How are evaluation designs different from models (or plans) for data analysis?

2. List some problems that you may have come across in assigning values to responses an attitudinal scales.

3. What are some questions you will like to see answered in regard to the analysis of naturalistic qualitative data?

CHAPTER XV

BASIC STATISTICS FOR EVALUATORS OF DEVELOPMENT TRAINING PROGRAMS

Evaluators of development training programs should be familiar with some basic statistical concepts and techniques which they can put to work in the analysis of quantitative data. One would, of course, need years of rigorous training to become a professional statistician. However, the basic essential concepts and techniques of statistics can be learned much more quickly. Acquisition of arithmetical skills of high school level is perhaps the only pre-requisite for acquiring functional *statistical literacy*.

What is statistics? Statistics is the science that deals with the collection, tabulation and systematic classification of quantitative data (and of qualitative data where it can be translated into quantitative terms). There is *descriptive* statistics and there is *inferential* statistics. Descriptive statistics helps us describe in numerical terms entities, proportions, characteristics and conditions that we see in real life. Inferential statistics helps us in using occurrences of events in inductive thinking; and to infer the general from the specific.

Statistical Needs of Trainers-Evaluators

On the basis of our experience within the context of DSE workshops on the evaluation of basic education and development training programs, we have identified the following as basic statistical needs of development trainers.

1. Summarizing the scores of members of a group to understand the characteristics and tendencies of the group as a whole. This means learning

to determine the mode, median, mean, standard deviation (and variance) of a group of scores; in other words learning to measure the central tendency of a group of scores.

2. Testing for significant differences between and among the performances of different groups, or differences between the same group at two different times.

3. Testing for correlations, or for the going together of two particular characteristics in the same individual or group. (An example of correlation would be the going together of literacy of parents and social mobility of their children in the Third World.)

In the following, we will present the statistical concepts and techniques required for fulfilling the needs of evaluators of development training programs. We present these techniques as recipes, with emphasis on the procedural steps that must be taken in their application. Nothing at all will be said about the mathematical logic of the statistical formulas chosen for discussion. Those interested in developing deeper understandings of these statistical concepts and techniques should refer to any standard work on statistics for education and psychology.

Measures of Central Tendency

Measures of central tendency are group measures. They help us make statements about the characteristics of a group as a whole. These are:

1. Mode
2. Median
3. Mean
4. Standard Deviation (and Variance)

The mode is that score that occurs most frequently in a group of scores. Examine the test scores of a group of ten trainees given below:

Trainees	Scores on a test on knowledge of nutrition (x)
1	80
2	70
3	73
4	72
5	78
6	60
7	63
8	72
9	73
10	72

Note that (x) is the algebraic sign for raw scores.

While visual inspection may help us find the mode (also the median or the mean) of small groups of scores, evaluators typically develop a frequency table first, with scores ranked in an ascending or a descending order:

Score (x)	Frequency (f)
80	1
78	1
73	2
72	3
70	1
63	1
60	1

An examination of the frequency table above indicates that 72 is the most frequently occurring score in this group of scores. Thus, the mode of the group is 72.

The frequency table above also tells us that the range of scores is between 60-80. This can be a useful piece of information in the process of data analysis.

The median is a score that divides a group of scores into two equal halves. Thus, 50 per cent of the scores in a group are above the median score and 50 per cent are below it.

In the frequency table above, we find that our 5th student is opposite the score of 72, and that there are three students with that score. This is how we proceed in such cases to find out the median. All the three scores can be seen to be located between 71.5 and 72.5. Since there are three scores, each of the three can be seen to add an increment of .33 (72.5-71.5 divided by 3) to 71.5 as we move from 71.5 to 72.5. We need one increment of .33 to locate the invisible mid-point. Hence the median of the group of ten scores above is 71.5 + .33 = 71.83.

We should indicate here that both the mode and the median are sterile statistics. There is nothing much we can do with them, after we have got them.

The mean (or the average) of a group of scores, however, provides a useful statistic for measuring the central tendency of groups. Examine the following:

(x)	(f)	f (x)
80	1	80
78	1	78
73	2	146
72	3	216
70	1	70
63	1	63
60	1	60
	Total	713

$$\text{Mean } (\bar{X}) = \frac{713}{10} = 71.3$$

Standard Deviation (and Variance)

Variance is a statistical measure of how individual scores in a group vary from the mean of the group. The algebraic sign for variance is s^2, where s stands for standard deviation.

The mean by itself may not be an adequate measure of the central tendency of a group. One may have to work out the standard deviation (or the variance) of a group to describe it more adequately.

Examine the two groups below:

Group A		Group B	
x_1	f_1	x_2	f_2
80	1	85	1
78	1	83	1
73	2	78	2
72	3	77	1
70	1	67	2
63	1	65	1
60	1	58	1
		55	1

Mean $\bar{X}_1 = 71.3$ Mean $\bar{X}_2 = 71.3$

These two groups have the exact same mean but even a visual inspection tells us that the two groups are different. Their modes differ (72 versus 67 and 78). Their medians differ (71.83 versus 67). But most importantly the scores among each group vary differently from each other. That is, the two groups have different variances (or standard deviations).

How to Measure Standard Deviation

Standard deviation (s) can be calculated by using the following formula:

$$s = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}}$$

Where X stands for the various scores in the group of scores

n stands for the number of scores, and

Σ stands for summation.

Now the standard deviations of the two groups can be worked out as follows:

Group A

$$\Sigma X^2 = 80^2 + 78^2 + 73^2 + 73^2 + 72^2 + 72^2 + 72^2 + 70^2 + 63^2 + 60^2$$

$$= 6400 + 6084 + 5329 + 5329 + 5184 + 5184 + 5184 + 4900 + 3969 + 3600$$

$$= 51,163$$

$$(\Sigma X)^2 = (80 + 78 + 73 + 73 + 72 + 72 + 72 + 70 + 63 + 60)^2$$

$$= (713)^2$$

$$= 508,309$$

$$s = \sqrt{\frac{51163 - \frac{508,309}{10}}{10-1}}$$

$$= \sqrt{\frac{51163 - 50836.9}{9}} = \sqrt{\frac{326.1}{9}}$$

$$= \sqrt{36.233} = 6.0194$$

Group B

For Group B

$$\Sigma X^2 = 51803$$

$$(\Sigma X)^2 = 508,369$$

$$s = \sqrt{\frac{966.1}{9}} = \sqrt{107.344}$$

$$= 10.3607$$

The two standard deviations, 6.0194 for Group A and 10.3607 for Group B tell us that the two groups are quite a bit different from each other in their variance.

The standard deviation (s) and variance (s^2) are useful statistics and can be put to important uses.

The t-Test for Independent Samples

Independent samples are those which are drawn independently from a population without any pairing or matching or any other relationship existing between the members of the two groups. (Ideally, these samples should be drawn randomly but it may not always be possible to do so in field situations.)

To see if two independent samples of trainees are statistically different from each other in regard to their performance on a knowledge test for example, a t-test can be used. The t-test for independent samples is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{\Sigma x_1^2 + \Sigma x_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where $\bar{X}_1$ is the mean of scores in Group 1

$\bar{X}_2$ is the mean of scores in Group 2

Σx_1^2 is the sum of the squared deviations of scores from the mean in Group 1

Σx_2^2 is the sum of the squared deviations of scores from the mean in Group 2

n_1 is the number of scores in Group 1

and n_2 is the number of scores in Group 2.

One does not have to memorize this formula. We only need to know when the t-test of independent samples should be used and where to find it.

Let us examine the statistics for two independent samples below:[1]

Group 1	Group 2
$\bar{X}_1 = 14$	$\bar{X}_2 = 10$
$\Sigma x_1^2 = 106$	$\Sigma x_2^2 = 44$
$n_1 = 15$	$n_2 = 15$

Then,

$$t = \frac{14 - 10}{\sqrt{\left(\frac{106 + 44}{15 + 15 - 2}\right)\left(\frac{1}{15} + \frac{1}{15}\right)}}$$

$$= \frac{4}{0.84} = 4.76$$

The degrees of freedom for the t-test of independent samples is $n_1 + n_2 - 2$ which works out to be 28 in this case. Now look at the table of t values in the first table at the end of this chapter. With 28 degrees of freedom, a t value of 3.674 is significant at .001 level of confidence. In other words, with a t-value of 3.674, there is only 1 chance in a thousand that the difference perceived between the two groups is by mere chance and is not a real difference. With our t value of 4.76, we can confidently claim a real difference between the two groups.

The t-Test for Nonindependent Samples

When two groups of trainees are matched on some characteristic or when before and after scores of the same group are being used, we are dealing with nonindependent samples. The formula for finding the t value for matched or before and after groups of scores is:

---

[1] This example is condensed from one given in Donald Ary, Lucy Chester Jacobs and Asghar Razavieh, Introduction to Research in Education, Holt, 1979.

$$t = \frac{\bar{D}}{\sqrt{\dfrac{\Sigma D^2 - \dfrac{(\Sigma D)^2}{N}}{N(N-1)}}}$$

where D is the difference between the paired scores of two matched individuals or the before and after scores of the same one individual

$\bar{D}$ is the mean of the differences (D's)

ΣD^2 is the sum of squared difference scores, and

N is the number of pairs.

The degrees of freedom in the case of a t-test for nonindependent samples is N-1. The same table of t values should be consulted to determine the statistical significance of the t value obtained.

Analysis of Variance: The F-Test

If comparisons between more than two groups are involved, a statistical procedure called the analysis of variance (ANOVA) and the F-test can be used. Readers are referred to a standard text on statistics for a description of the ANOVA procedures.

The Chi-Square Test of Significance

Evaluators may sometimes have to compare not means of scores (as in the case of t-tests above) but proportions and frequencies. The questions may take such form: Is there a difference in terms of age for women coming to a health clinic? Do men and women differ in regard to their membership in cooperative societies? Is there a difference between urban and rural schools in terms of their adoption of educational innovations? Does the size of landholding make a difference in a farmer's use of new agricultural technology?

To answer these types of question the chi-square (χ^2) test can be used:

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

where f_o is the observed frequency of responses, and

f_e is the expected frequency - the frequency of responses that would have been theoretically expected, if there had been no differences between groups or entities being compared.

Let us look at the example below:

TESTING FREQUENCY DISTRIBUTION OF HIGH ADOPTER AND LOW ADOPTER SCHOOLS IN RELATION TO URBAN AND RURAL ENVIRONMENT USING CHI SQUARE[1]

	Environment		Total
	Urban	Rural	Frequency
High Adopter Schools	12 (8)	3 (7)	15
Low Adopter Schools	4 (8)	11 (7)	15
	16	14	30

The frequencies actually observed are 12, 4, 3 and 11. The expected frequencies have been shown in parentheses. If being urban did not make any difference in regard to innovative behavior, we should have expected high adopter and low adopter schools to be equally divided in the urban area and we should have obtained the same frequency of responses ($\frac{12+4}{2} = 8$). Similarly, if being rural did not make any difference in regard to innovative behavior we should have expected to see the same number of high and low adopting schools in the rural areas ($\frac{3+11}{2} = 7$).

---

[1] Orest Paul Ochitwa, A Study of the Organizational Climate of High and and Low Adopter Elementary Schools in the Province of Saskatchewan, Canada. Doctoral Dissertation, Indiana University, 1973.

It should be noted that determining expected frequencies is not always a matter of averaging across categories. Expected frequencies are theoretically determined. Suppose there are 30 per cent high school graduates among your literacy teachers and the remaining 70 percent have lower than high school education. Suppose we ask all our literacy teachers a question about their perceptions in regard to their needs for further training and want to find out if responses differ on the basis of teacher's academic qualification: high school versus lower than high school. The expected frequencies in this case will be proportionate to the 30:70 division in the population. (It will not be a 50:50 division.)

Now to answer the question if real differences exist between urban and rural schools in regard to innovative behavior, we will apply the chi-square test:

$$x^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

$$x^2 = \frac{(12-8)^2}{8} + \frac{(4-8)^2}{8} + \frac{(3-7)^2}{7} + \frac{(11-7)^2}{7} = 2 + 2 + 2.3 + 2.3$$

$$= 8.6$$

The formula for working out the degrees of freedom in the case of x^2 is (C-1)(R-1). C is the number of categories in the column; and R is the number of categories in the row. Thus, the degrees of freedom in our case are (2-1)(2-1) = 1.

Let us now examine the table on χ^2 values. (See Table 2 at the end of this chapter.) With 1 degree of freedom we need a χ^2 value of 6.635 to assert that a difference exists between two sets of responses, with a .01 level of confidence (that is, there being only 1 chance in a 100 for the difference to exist by mere chance and not be a real difference). Our χ^2 value is 8.6 which is even higher than 6.635; and hence we can assert that there is a difference between urban and rural schools in regard to their innovativeness, with .01 level of confidence (also expressed as $p = < .01$).

Spearman Rank Correlation Coefficient

Correlation means whether two properties or characteristics of individuals go together. Or, to put it differently whether two particular characteristics covary.

Take the example of a class that takes two tests, one for Reading and another for Mathematics. The question can be asked: Are students who are good in Reading also good in Mathematics and vice versa. Or, are the student scores in Reading and Mathematics correlated?

The formula for determining correlation coefficient is

$$\rho(\text{rho}) = \frac{1-6\Sigma d^2}{n(n^2-1)}$$

where d is the difference in the ranks in two scores (Reading and Mathematics) for the same individual and n is the number of cases involved.

Let us examine the following data.[1]

---

[1] From H. S. Bhola, Evaluating Functional Literacy. Amersham, Bucks, U.K.: Hulton Educational Publications, 1979.

Student	Maths Score	Maths Rank[1]	Reading Score	Reading Rank	d	d^2
A.B.	22	1	53	1	0	0
D.G.	20	2	32	4·5	2·5	6·25
C.G.	17	3	48	2	1·0	1·00
I.F.	15	4·5	36	3	1·5	2·25
H.H.	15	4·5	22	8	3·5	12·25
M.B.	14	6	30	6	0	0
S.R.	11	7	28	7	0	0
T.F.	8	8	32	4·5	3·5	12·25
G.K.	7	9	18	9	0	0
P.S.	4	10	15	10	0	0
						$\Sigma d^2 = 35{\cdot}00$

$$\begin{aligned}\rho &= 1 - \frac{6\,(35)}{10\,(100-1)}\\ &= 1 - \frac{210}{990}\\ &= 1 - 0{\cdot}21\\ \rho &= 0{\cdot}79\end{aligned}$$

Let us note that correlations can be both positive and negative. They can vary from -1 to +1. Correlations between 0.00 and 0.30 are considered low; those between 0.30 to 0.60 are considered medium; and those between 0.60 to 1.00 are considered high.

We must also remember that correlations do not establish cause and effect relationships. If scores in Reading and Mathematics are highly correlated, it does not necessarily mean that high scores in Reading cause high scores in Mathematics.

---

[1] When more than one of the same scores occur, the midpoint between the rank preceding and the rank following is assigned to all tied scores.

Summary

Evaluators must all have a functional level of statistical literacy. With very little effort on their part, trainers-evaluators can learn to measure means and variances of groups of scores; apply t-tests and χ^2 tests to test for real differences between the responses made by groups; and can work out correlations between two particular characteristics of an individual or a group.

Things to Do or Think About

1. Do you have in the institute's library any of the standard books on statistics that could be used by educators and development workers generally? If not, it is time to order some books on statistics for your institute.
2. Look at some of the test results already in your files and work out means and variances for various groups of scores.
3. If you have some before and after test data for a group of trainees, apply to it the appropriate t-test.
4. Are knowledge of nutrition scores always correlated with better nutritional habits? Obtain a set of scores for knowledge and for behavior and work out the correlation coefficient.

Table of *t* Values

Degrees of freedom	Probability			
	.1	.05	.01	.001
1	6.314	12.706	63.657	636.619
2	2.920	4.303	9.925	31.598
3	2.353	3.182	5.841	12.924
4	2.132	2.776	4.604	8.610
5	2.015	2.571	4.032	6.869
6	1.943	2.447	3.707	5.959
7	1.895	2.365	3.499	5.408
8	1.860	2.306	3.355	5.041
9	1.833	2.262	3.250	4.781
10	1.812	2.228	3.169	4.587
11	1.796	2.201	3.106	4.437
12	1.782	2.179	3.055	4.318
13	1.771	2.160	3.012	4.221
14	1.761	2.145	2.977	4.140
15	1.753	2.131	2.947	4.073
16	1.746	2.120	2.921	4.015
17	1.740	2.110	2.898	3.965
18	1.734	2.101	2.878	3.922
19	1.729	2.093	2.861	3.883
20	1.725	2.086	2.845	3.850
21	1.721	2.080	2.831	3.819
22	1.717	2.074	2.819	3.792
23	1.714	2.069	2.807	3.767
24	1.711	2.064	2.797	3.745
25	1.708	2.060	2.787	3.725
26	1.706	2.056	2.779	3.707
27	1.703	2.052	2.771	3.690
28	1.701	2.048	2.763	3.674
29	1.699	2.045	2.756	3.659
30	1.697	2.042	2.750	3.646
40	1.684	2.021	2.704	3.551
60	1.671	2.000	2.660	3.460
120	1.658	1.980	2.617	3.373
∞	1.645	1.960	2.576	3.291

SOURCE: Abridged from Table II in R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research* (New York: Hafner, 1974).

BEST COPY AVAILABLE

Table of χ^2

Degrees of Freedom	P = .99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01
1	.000157	.000628	.00393	.0158	.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892

SOURCE: R. A. Fisher, *Statistical Methods for Research Workers*, 14th ed. (New York: Hafner, 1973), Table III

BEST COPY AVAILABLE

CHAPTER XVI

WRITING EVALUATION REPORTS

An evaluation study to be most useful must end in a written report. A written report serves at least two purposes. First, it provides an opportunity to the evaluator to organize the data collected, to systematize thinking, to draw conclusions, and to weigh and consider the implications of the study as well as its limitations. Not to write a report of an evaluation study, after data gathering and data collation has been done, would be like grazing a cow in the pastures all day and not milking it in the evening! Second, the evaluation report serves as the instrument of communication between and among professional colleagues and others interested in the same or similar problems and issues.

Evaluation studies have quite often been published; and, sometimes, have brought high professional rewards to evaluators. However, publication and rewards of fame and fortune are not the right expectations to have when writing an evaluation report. These rewards may come, but one should not strain to get them everytime one sits down to write an evaluation report. It is much more realistic to think in terms of making a few copies of the evaluation report to be shared, first and foremost, with professional colleagues who should know what your evaluation study has found; who can discuss your conclusions and suggestions with you, and who, perhaps, can use the report to improve their performance within the setting of the institution to which you all belong.

We like to make a distinction here between a basic professional report and other written or oral presentations. The evaluator should prepare one basic and comprehensive report on the evaluation study. This basic report then should be used to make different written and oral presentations to different groups of people who may be interested, among them, policy makers and planners, politicians, extension workers, and even farmers and housewives who are often the subjects of our developmental efforts.

The Essential Objective of the Basic Report and Its Parts[1]

The essential objective of an evaluation report is to make a complete record of an evaluation experience including the background and the context of the evaluation questions; the assumptions made in posing the question(s); the evaluation design and tools used in data collection; the results obtained; conclusions drawn; and practical implications developed from the conclusions of the evaluation study. In other words, the evaluation report is a sort of a mirror image of an evaluation proposal as discussed in Chapter VI "Writing Formal Proposals for Evaluation Studies."

An evaluation report, however, is more than an evaluation proposal written in the past tense. A good evaluation report includes all the information necessary for a reader to be able to evaluate the evaluation study itself. That is, the reader should know exactly what was done and how; using what samples and what questions; and what data was actually collected. The reader must also be told of the structure of the argument used in data collation and analysis, and what conclusions were drawn and why. In all cases, the reader should thus be able to see the strengths of the study as well as its limitations; and, where necessary, the reader should be able to do a "secondary analysis" of the data on his or her own to draw independent and even alternative conclusions. This means that actual tools and instruments, and any specimens of stimulus materials used in the study, should become part of the report as "Appendices."

---

[1] Remarks on report writing in this section of the chapter apply mainly to evaluation studies that use the scientific paradigm.

This does not mean, however, that all raw data should become part of the report or should be put in the appendices. A report is not a device for storing and filing all the raw data that was collected for an evaluation study. Data included in the report or in the appendices should be in collated form, already organized into tables and displays. In some cases, it may be necessary to present data in sufficiently "disaggregated" form so that it is possible for the reader to aggregate data in different ways to test assumptions and conclusions of the original evaluator; and, as we have suggested above, to draw alternative conclusions.

An evaluation report should typically (but may not always) have the parts and sections discussed below:

The title page

The title page of the report should show the title of the evaluation study, the name of the evaluator(s), the institutional affiliation of the evaluator(s), and the date when the report was issued.

The title given to the report should faithfully reflect the purpose and scope of the evaluation study. This same exact title should then be used throughout the study without arbitrary variations. In some cases, it may be useful to have both a long title and a short title for the same study. Once chosen, these titles should be used in other parts of the report without change. The date of issue of the report should be shown on the title page, as we have suggested earlier, but somewhere in the body of the report one should also indicate the dates and period of time during which data was actually collected. (It is possible to collect data in the first six months of 1982 and publish a report in 1984.)

The abstract

A one- to two-page abstract (that is, of about 500 words) should precede the evaluation report. This should be a complete summary and must include information about the evaluation question; samples and procedures used; findings and their program implications. A person who does not read the full report should yet get a fairly good idea of contents of the study from reading this abstract.

General background

The first part in the main body of the evaluation report should be the general background of the study. This material will not have to be written anew, but should be adapted from the evaluation proposal written earlier. Put training for development in a larger perspective of human resource development for social change. Comment on the need for evaluation of training, in general. Be brief. No more than a page or two should be utilized.

Focus on your development sector and institution

Focus should then shift to your specific development sector such as agriculture, cooperatives, health extension, nutrition or family planning, and to your institution. Talk about the role of your training institution and its contribution to the training of needed manpower for development. Once, again, brevity is important. One or two pages of tightly written material should be enough.

The training model in use

Present the bare-bones of your model of training. Answer questions such as: What are the assumptions made about the change process in the training model in use? What are the assumptions made about the change agent's role? What are the objectives of training? What are the special training methods used? What are the K-A-P (Knowledge-Attitude-Performance) claims being made in behalf of the training program? (All these questions will not have to be answered in each and every evaluation report. Nor will these questions be answered in the order in which they have been listed here. These are the questions to "think with" as evaluators sit down to write their final reports.)

The evaluation question(s) asked

The evaluation questions as answered in the evaluation study should be carefully listed. This list of questions must later be used in the collation and analysis of data. These questions should have two linkages. One, these questions should relate with the training model-in-use discussed earlier. It should be clear how the training model-in-use generated that set of questions. Two, these questions should be linked with the subsequent organization of data in a later section and should provide the organizing principles for data analysis.

Why was this feedback necessary

This is in fact a justification for the choice of particular evaluation questions from a whole array of possible questions generated by the training model and the institutional needs for feedback. The material from the earlier evaluation proposal on "justification" and "significance" should be used for developing this section.

Assumptions made

Assumptions made, general to the change and training processes, and specific to the institutional and field settings of your evaluation study, may be stated here, as relevant. Some of these assumptions will have been stated in the earlier proposals. Some others may have been uncovered during the process of implementation.

Procedures and methods used

This section should include the general evaluation design; analyses of concepts used and special definitions assigned to terms; indicators used and the process of their development and choice; criteria to be used for evaluating success or failure of the program; samples chosen (and as originally intended); tools and instruments used (which must be placed in the appendix); field work procedures followed, including recruitment and training of investigators and time and duration of the field work phase.

Evaluation design. Go back to the evaluation proposal and reproduce, with adaptations if any were made, the evaluation design used in the study. This model will now be used as a model of data analysis. (See the discussion on models of data analysis in Chapter XIV of this document.)

Conceptual analysis and definitions. You may have undertaken conceptual analyses of some concepts such as humanism and self-reliance; or may have given special definitions of your own to such words as dropout, literate, etc. These should be included in this section.

Indicators -- their development and choice. The process used in going from larger categories such as self-reliance to subcategories of larger concepts, and, finally, to the choice of indicators which can be measured, should be clarified in this section of the evaluation report.

Standards and criteria of success. It is important to indicate in the report, the levels of expectations and standards set for judging success or failure of the program being evaluated. The reader should have an idea about whether to be satisfied or dissatisfied with the 30 per cent dropout rate from a literacy class or the 10 per cent rate of success in the rehabilitation of the handicapped.

Samples and units of response. Explain sampling procedures. Define the samples that were planned to be used and the samples that were actually used. What were the units of response? Was it the housewife, or was it anyone else (husband, an older child), speaking in behalf of the family? Was it the chairman of the committee being interviewed, or was it anyone in the committee (or more than one person taking turns), speaking in behalf of the committee? What was intended? What actually happened?

Tools and instruments. The variety of tools and instruments used should be indicated and their choice justified. Any special procedures used in developing and pre-testing tools should be given. Changes made in tools and instruments on the basis of pretesting should be highlighted. The tools and instruments should be included in the appendices.

Field work. This section should clarify any strategy implicit in the field work -- coping with distances, or with weather conditions; piggybacking on existing systems of transportations and supervision, etc. It should, additionally, include a description of field work procedures and experience. Were investigators used? How were they trained? How were they supervised? How was communication between the evaluator and the field investigators maintained? Was there a small pilot study conducted before the final study? Did some data have to be collected twice? What was the time and duration of the field work? Was it found necessary to use a follow up questionnaire or interview to supplement the original data?

Limitations and breakdowns. This section should look backward on field work experience, and forward to the section on data analysis and should indicate any breakdown that occurred in field work and any limitations that became apparent in data collation and analysis later.

Recording of findings

This section is the heart of any evaluation report. It has to present all relevant data in collated form, in effective displays of tables, charts, graphics, lists and vignettes to serve as evidence for all answers given, comments offered and conclusions drawn. The list of questions drawn up earlier and the model of analysis discussed before should be used to organize the collation of data, its display, analysis and interpretation.

A separate section may deal with questions not originally asked but which the available data was able to answer.

Discussion of results

The findings must be discussed in regard to the implications for action, and guidelines for future training design. The evaluation results obtained must be discussed in terms of expectations, standards and norms. These should also be discussed in terms of the strength of data, correlations and possible causal links.

Further evaluation and feedback needs

Karl Popper has said that our knowledge and ignorance increase together! A successful evaluation study, by creating new information, might also tell us what we are ignorant of, or need to know more about. New feedback needs should be identified and suggestions in regard to further evaluation studies should be made.

Bibliography

Make a list of books, documents and government reports used in the implementation of the evaluation study and in writing the report. In case of official documents, indicate if they are available to the public and where those might be obtained or consulted.

Appendices

The following kinds of items should go in the appendices: copies of tools and instruments; specimens and exhibits where appropriate; collated data not used in the body of the report but of interest to readers and evaluators; lists of names of people, and institutions that cooperated with the evaluator(s) in the conduct of the study; field work schedules, maps, etc.

Writing Reports of Naturalistic Evaluation Studies

Report of a naturalistic evaluation study will typically be in the form of a Case Study. Some of the ideas discussed above in regard to the writing of reports of evaluation studies using the scientific paradigm may have some implications for writing reports of naturalistic evaluation studies as well. It must be understood, however, that case study is a special genre of writing, with its own form and style. Unfortunately, there are no rules for writing case studies: one learns by reading case studies written by others, and then by writing one's own.

While it is not possible to suggest rules for writing case studies, it is possible to make some fruitful comments on the general nature of case studies. For example, it is useful to compare the writing of case studies with creative writing. Creative writing has been described by literary critics as the "imaginative reconstruction of reality." By analogy, case study writing can be described as the "theoretical reconstruction of reality, imaginatively done." Thus, case study is the child of both theory and imagination. A case study defines, categorizes, compares, questions, hypothesizes, and explains by reconstructing reality in theoretical terms. At the same time, case study provides "thick descriptions" of environments, actors, events, and consequences giving the reader a feeling of involvement and a sense of having been there. Thus it is that case study is a theoretical reconstruction of reality, imaginatively done.

Reports to Nonspecialists:
Written and Oral Reports

As we have suggested earlier, trainers-evaluators should begin with writing one basic professional report. This basic report should then be used for writing short written and oral reports for the nonspecialist. These reports should be written to suit the special interests of the audience to whom the report is addressed. Oral reports should be made both informative and interesting, using appropriate audio-visual aids. These written and oral reports to special audience should in fact become part of the process of dissemination of evaluation results.

Summary

The evaluator must write a basic professional report of the evaluation study completed. This basic report should then be used to develop a variety of written and oral reports to suit the needs and interests of various nonspecialist audiences. The basic professional report of an evaluation study can be the mirror image of the proposal of the evaluation study developed earlier as part of planning. The report of the naturalistic evaluation is typically a case study. It offers special challenges and, at its best, is the theoretical reconstruction of reality, imaginatively done.

Thinks to Do or Think About

1. Examine the report of an evaluation study recently completed at your center or at another development training institute. Do you find it to be a complete and comprehensive report? How would you outline the report to make an improvement on the present version?

2. Have you read a case study recently? Did you "enjoy" reading it? Did you "learn" important professional ideas and insights from the case study?

3. Prepare an oral presentation for a group of farmers based on an evaluation study done in your country on the subject of agricultural innovation.

CHAPTER XVII

POLITICS OF EVALUATION AND EVALUATION STANDARDS

We will end this monograph on Evaluating Development Training Programs with a discussion of the politics of evaluation; and a listing of the standards of evaluation for an evaluation of the evaluation process itself.

Handling the Politics of Evaluation

Information is power. Information can be put to political uses. Hence, evaluation, which creates somewhat objective information on the effectiveness of development training programs or on the impact of development actions, has political implications.

How to handle the politics of evaluation? No sure-fire formulas can be taught. In any case, most of us who have worked (and survived) within bureaucracies are not all that naive about the politics of survival and advancement within bureaucracies. Each one of us is perhaps somewhat qualified already in the art of "filemanship" and even "one-up-man-ship!" Yet, some general suggestions for handling the politics of evaluation may be in order.

There are two aspects to the politics of evaluation: (a) the evaluator should not get punished for doing the evaluation which may be seen as having produced "embarrassing" information; and (b) the information produced by the evaluation study should be put to practical use. Political problems do arise when, on the one hand, the evaluator seeks to make too much capital out of the evaluation study; and, on the other hand, creates information that threatens the various stakeholders within the system. Without compromising one's personal and professional integrity, one can do things, however, which will cool the politics surrounding the evaluation study.

Defend your right to undertake evaluation

Defend your right to conduct the evaluation. Let people know that evaluating is an integral part of good training. Quote from a presidential speech, planning documents, or from published prospectuses or reports of the training center or institution. Your institution is bound to have declared evaluation to be a necessary part of its mission, though no one may have paid much attention to this particular objective. In an educational setting (as distinguished from an administrative setting), the right to evaluate can be defended as part of your research interest. You, as a teacher, are supposed to have a research interest.

Keep a low profile

There is need for an evaluator to keep a low profile and have a sense of modesty about the evaluation study done. The evaluator should not demand to be considered a star in the institutional horizon. The report should be presented without too much fanfare, as a matter-of-fact collection of feedback information on the program. It should not be touted as a breakthrough of some sort.

Framework of expectations for evaluation results

No program will ever be found to be performing at 100% efficiency level. Especially in social change programs, participation levels of 30% may sometimes be deemed satisfactory. Before presenting the feedback on performance of a program, one must indicate what would be a reasonable level of expectation of performance. Findings should then be presented within such a framework. In other words, the readers of an evaluation report should be provided with standards and yardsticks with which to judge the success or failure of a training program or a development action. Without norms, readers may not know whether to be satisfied or to be dissatisfied with a particular set of results.

As I have said elsewhere, the focus should be on finding causes, not culprits. This is not to say that the program staff is never at fault and that as evaluators we should be finding alibis for them. Yet, processes and personnel must not be confused in the allocations of credit and blame. Things must be kept in balance.

Begin with a "draft" report

An important part of the political strategy may be to present the evaluation report to colleagues in a "draft" form, offering to do a final draft on the basis of colleagial discussion and review. In a revision that follows, it will be important to neutralize the politics but *without* compromising the integrity of results.

Indicate possible actions

Indicate the actions that must be taken to make use of the findings of the study. Distinguish between things within the institution's control and those outside its control. Start with what the institution can do within its existing mandate -- such as curriculum revision, preparation of new testing procedures, etc. If the implementation of findings demands additional work, offer to do it singly, or with the help of a group or a committee.

What we have suggested above may not always work, but it will increase the chances of an evaluation study influencing actions within the setting of a training center or a training institute.

Evaluation Standards:

Evaluation of Evaluations

Evaluators should themselves be held accountable. Their work must be judged according to some agreed upon standards of technical competence and ethics.

The Joint Committee on Standards for Educational Evaluations[1] of U.S.A. has developed 30 standards which the committee suggests should become the working philosophy of evaluators and should guide and govern the evaluation efforts of educators (and development workers). A summary of these standards is provided below:

Summary of the Standards for Evaluations

A. Utility Standards

Evaluation should serve practical information needs.

1 A_1 Audience Identification

Audiences involved in or affected by evaluation should be identified.

2 A_2 Evaluator Credibility

Evaluator should be both trustworthy and competent.

3 A_3 Information Scope and Selection

The scope and selection of information collected should enable pertinent questions to be answered.

4 A_4 Valuational Interpretation

Value judgments used by evaluators should be made clear to readers.

5 A_5 Report Clarity

Objectives, procedures used, findings, and recommendations should be clearly stated.

---

[1] The Joint Committee on Standards for Educational Evaluation, *Standards for Evaluations of Educational Programs, Projects and Materials*. New York, N.Y.: McGraw-Hill Book Co., 1981.

6 A_6 Report Dissemination

Findings must be disseminated for use.

7 A_7 Report Timeliness

Evaluation must be completed on time for use by decision makers.

8 A_8 Evaluation Impact

Evaluators should encourage follow through by the concerned audiences.

B. Feasibility Standards

Evaluation should be realistic, prudent, diplomatic and frugal.

9 B_1 Practical Procedures

Procedures should be practical and should avoid disruptions of normal work.

10 B_2 Political Viability

Evaluators should attract cooperation of various interest groups, avoid their attacks, ensure against misuse of results.

11 B_3 Cost Effectiveness

Results should justify resources expended.

C. Propriety Standards

Evaluation should be conducted legally and ethically and should contribute to human welfare.

12 C_1 Formal Obligation

Formal obligations and contracts may be developed between various parties involved (especially in the case of external evaluations).

13 C_2 Conflict of Interest

Should be avoided and where unavoidable should be dealt with openly and honestly.

14 C_3 Full and Frank Disclosure

Pertinent findings should be fully disclosed; limitations should be frankly stated.

15 C_4 Public's Right to Know

Public's right to know of evaluation results should be respected (unless it is clearly a matter of individual privacy or public safety).

16 C_5 Right of Human Subjects

Rights of human subjects should be respected and protected.

17 C_6 Human Interactions

In their interactions with subjects, evaluators should respect dignity and worth of individuals.

18 C_7 Balanced Reporting

The reporting should balance both strengths and weaknesses of what is evaluated.

19 C_8 Fiscal Responsibility

Financial and other resources spent should be accounted for.

D. Accuracy Standards

Evaluation should convey technically adequate information.

20 D_1 Object Identification

What is being evaluated should be clearly identified.

21 D_2 Context Analysis

Context of evaluation should be sufficiently described so that its influences on the object evaluated can be identified.

22 D_3 Describe Purposes and Procedures

The purposes and procedures of evaluation should be described in enough detail.

23 D_4 Defensible Information Sources

The sources of information should be described so reader can see if those are defensible sources.

24 D_5 Valid Measurement

Evaluation instruments should be constructed and applied in ways to ensure validity.

25 D_6 Reliable Measurement

Evaluation instruments should be constructed and applied in ways to ensure reliability.

26 D_7 Systematic Data Control

Data should be reviewed and corrected at various stages of the study.

27 D_8 Analysis of Quantitative Information

Should be appropriate and systematic.

28 D_9 Analysis of Qualitative Information

Should be appropriate and systematic.

29 D_{10} Justified Conclusions

Conclusions should be explicitly justified.

30 D_{11} Objective Reporting

The reporting should be objective and unbiased.

Some of these standards may seem too tough and some too squeamish and overly fastidious to evaluators working in cultures other than the United States where these standards were developed. Evaluators everywhere should, however, take these standards into account to the extent feasible.

Concluding Comments

The understanding has now emerged that the quality of populations is an important determinant of development. Trainers in development training institutions and centers prepare change agents -- literacy workers, adult educators, health technicians, nutritionists, cooperators and agriculture extension workers. These change agents must, in turn, teach housewives, train farmers and upgrade factory workers to improve the quality of populations within developing societies. Thus, the work of development trainers is an important aspect of the overall development effort.

Indeed, the development training sector is large in many development countries and significant in all of them. Scope for improvement both in training design and the evaluation of training is immense. Evaluation, especially, can bring high economic and social returns. This book has tried to provide trainers-evaluators in the various development sectors some useful ideas both in training design and the evaluation of training.

There are a few implicit messages in this book that should be made more explicit. Training design has become a speciality, but it is not something that only supermen do or can do. Evaluation of the effectiveness of development training, again, has become systematic and technical, but, again, it is not something beyond our reach; something which only the trained university research can do.

In basic terms, the ideas of training design and training evaluation are quite simple. They are not beyond our intellectual capacity, nor beyond our personal and institutional resources. These tasks can be performed now, without much outside help, as part of our daily duties, and using the already available institutional resources. Needed is a personal commitment to grow as a professional worker and to do the best possible job within the circumstances. We challenge our readers to make such commitments.

Summary

Evaluation produces information which can be used politically. Evaluators must learn to manage the politics of evaluation so that they do not get punished for "speaking truth to power"; and at the same time the information they produce is put to practical uses by decision makers. Thirty standards for judging the performance of evaluators themselves have been listed.

Things to Do or Think About

1. What do you think of the practicality of suggestions made in the first part of this chapter for managing the politics of evaluation?
2. Evaluate a recent evaluation study in terms of the 30 standards for evaluation listed above.

APPENDIX "A"

A NOTE ON THE ACTION TRAINING MODEL (ATM)

The Action Training Model (ATM) is an in-service training model. It is so called because it demands action from trainees in the application of skills learned during training in their own work on return to their institutions. Skills learned during training must be put to work as part of the daily duties back home, without loud and formal announcements of special projects and programs.

ATM has been developed under the aegis of the German Foundation for International Development, Education and Science Branch, as a result of a collaboration between the Foundation's Dr. Josef Muller and Professor H. S. Bhola of Indiana University, Bloomington, Indiana, a collaboration that began in 1976. Dr. John W. Ryan, then Director, International Institute for Adult Literacy Methods, Tehran, Iran, and now of the Literacy, Adult Education and Rural Development Department of Unesco, Paris, made important contributions to the earlier definition of the model as a participative training model. An action orientation has been added, over the years, to the earlier participation orientation. ATM is the result.[1]

Early beginnings: Mwanza workshop on literacy evaluation, 1976

The first definition of the model as a participative training model emerged within the context of a two-week workshop on the evaluation of literacy programs for middle-level literacy workers held in Mwanza, Tanzania during December 1976.

---

[1] Contributions of the Kenyan faculty (Joseph Dondo, Tom Mulusa, Alice Waka, and Daudi Nturibi) in the definition and particularly in the implementation of the Action Training Model are most gratefully acknowledged.

The emphasis at that time was on participative planning so that workshop organizers would not be merely creating opportunities for themselves to deliver lecture on topics they thought were important. The task was instead to design a workshop which the local participants considered important, at that particular stage in the implementation of their national literacy program. The planning was to be participative; and the organizers were to be among the participants, as equal as others in the group.

The following features of the Mwanza Workshop should be mentioned by way of an initial definition of the training model as it emerged in Mwanza:

1. The Mwanza workshop took place within the context of nonformal education in Tanzania. It was conducted under the aegis of the department of adult education. Since nonformal adult education is a less formalized sector of education, it may have provided a context for experimenting with different training methods and strategies.

2. Tanzania had an important on-going program in literacy at the time; evaluation of literacy was part of the concern of the government; and those who came to attend the workshop were actually responsible for evaluation within the Tanzanian literacy program. This meant that workshop participants knew quite clearly what their needs were. They did not have to invent evaluation problems; and the evaluation problems they did choose had relevance to their work.

3. The Mwanza workshop was designed as a "national" workshop. This had many advantages: it was, comparatively speaking, a low-cost workshop; more than a handful of middle level workers could be trained -- almost thirty in this case; and discussions at the workshop could be rooted within a real program and a particular social and political setting. (The Mwanza workshop

was attended by two observors from Kenya. On return home, they would help the Government of Kenya decide if Kenya needed and wanted a similar workshop on the evaluation of literacy and adult education programs.)

4. The workshop sought to train middle-level literacy workers who would themselves do most of the implementation of literacy evaluation as part of their day-to-day work.

5. It was a two-week workshop build around plenary sessions, group work and individual tutoring. Field visits to villages were not included in the program. This omission was deliberate because it was considered unnecessary to take rural development workers who spent practically all their working lives in the rural mileu back to visit villages to experience field realities!

6. The workshop was not planned as an "international" event attended by VIP's. It was a working workshop where participants labored. We stayed away from big hotels, and lived in a "retreat" setting, away from the city and without a number of cars and buses shuttling around.

7. The workshop came about at the end of long and systematic preparation. The aims, objectives, procedures and possible content of the workshop was thoroughly discussed. A five-page project description was prepared for use as a tool of communication between and among all concerned with the workshop.

8. The workshop was based on carefully written instructional materials. A monograph was completed specially for use in the workshop, namely, Evaluating functional literacy by H. S. Bhola. (This was later tested in other workshops on literacy evaluation and published as a book in 1979 by Hulton Educational Publications, Amersham, Bucks, U.K.). The monograph was accompanied by a dossier of instruments actually in use within literacy programs in various parts of the world.

9. The most important aspect of the workshop, however, was the participative planning of the workshop. The workshop was indeed "invented" in the local setting. There had been a lot of preparation for the workshop, but it was general not specific preparation. There were no prior decisions on what specific lectures will be given, by whom and at what time. The workshop preparation was in fact preparation for a whole set of possible workshops, only one of which would be required by the participants, within the context of the local needs of Tanzania at that particular time. Thus, preparation was comprehensive and open-ended; the workshop was anticipated but had to be re-invented, within the local setting, in participation with the learners. It should be mentioned here that the use of a written monograph for the workshop made the task easier. Participants were able to understand the part-whole relationship between their workshop and overall subject matter of evaluation. They knew from the written materials what a comprehensive elaboration of a subject or topic would be and how their own specific sharply defined needs related to the larger picture.

10. An important feature of the participative model was participation at the faculty level. One part of the objectives of the workshop was the capacitation of the local faculty through collaboration in teaching.[1]

11. There was considerable emphasis on group work, both for methodological and practical reasons. The groups provided an opportunity to all participants to discuss and assimilate important ideas and to be able to work on topics and skills of special concern to them in their particular work situations.

12. An innovative use was made of the wall space in the lecture hall. The

---

[1] At the Mwanza workshop we were fortunate to have the collaboration of Yusuf Kassam and E.P.R. Mbakile both of whom were experienced evaluators.

walls were plastered with paper. The learning needs as generated by individual participants in the planning session were written on those sheets. These sheets stayed there throughout the two-week period. The participants were able to see what they had wanted to see done, what had actually been accomplished, and what remained to be done. They always had a visual picture of the life of the workshop before them.

13. The workshop was steered by the participants through the mechanism of a steering committee. Every evening without fail, the steering committee met to review the experience of the day and on that basis to plan for the next day. In addition to the evaluative mechanism of the steering committee, the workshop was evaluated by sessions, by phases, and, summatively, at the end by Dr. Josef Muller and feedback provided to participants.

14. To reinforce the participative planning processes, daily programs and time tables were issued only at the end of the day -- to record what had actually happened rather than to have a time table to be strictly followed and by which the bell will ring. This strategy contributed both to the processes of formative evaluation and of participative planning of the workshop.

Second use of the model: Mombasa workshop on literacy evaluation 1977

Both the formative and summative evaluations of the Mwanza workshop were found to be excellent. The same participative model was therefore, followed in Mombasa, Kenya in 1977. The workshop was re-invented in Mombasa, using participative planning strategies and was, again, participatively implemented. All the various methodological features of the Mwanza model were retained;

yet in Mombasa we had a workshop that was new and unique to the needs of the Kenyan literacy program in 1977.

Toward the Action Training Model (ATM)

The participative model was good as far as it went. We were not sure, however, of what happened *after* the participants left for their homes. Had they been able to acquire enough information and skills in a two-week workshop to be able to do something with them as they returned to their posts? Had they been able to develop an evaluation proposal on their own and implement it? Did they get any support within the system to conduct an evaluation study, even if they knew what to do and how?

To increase the probability that participants will actually conduct evaluations in their day-to-day work to improve their performance, an Action Training Model was devised; to participation was added the commitment to action. The earlier model was augmented by three important features:

1. Actual development of evaluation proposals, dealing with some aspect of the trainee's own work, was made an essential part of the curriculum. Indeed, all evaluation concepts and skills were learned within the framework of preparing an evaluation proposal.
2. To give time to participants to actually implement their evaluation proposals, a long-term commitment was made to them and, in turn, a long-term commitment was demanded from them. The workshops were now organized as "sandwitches" -- a workshop was followed by a mid-term panel, which was, again, followed by another workshop. The same group of participants was supposed to go through the full cycle of about one year. This was an important program commitment made by DSE.

3. The capacitation of local faculty resources was made more systematic and there was a clear concern for institutionalization. For example, an Educational Evaluation Resources Committee was created in Kenya composed of local professional workers who would provide technical assistance to participants in the conduct of their evaluation studies during the one year cycle of their enrollment.

The beginning of the cycle A:
May 1979 evaluation workshop in Kenya

The ATM got its first trial in Kericho, Kenya in May 1979. Nineteen participants, trainers from centers and institutions in different development sectors (agriculture, cooperatives, health, nutrition, family planning), attended the workshop. A variety of evaluation proposals were completed during the workshop and participants went back to their workplaces with promises to work on their evaluation studies. They were to come back to a panel in January 1980, with evaluation data in hand, to learn what to do with that data. At the same time, an Educational Evaluation Resources Committee was established to help participants with their evaluation studies during the months of May 1979 to December 1979.

The experience of the EERC of working with participants was not all that satisfying and only six people showed up for the January 1980 panel. This was something of a shock to the organizers, though it need not have been. We know of no other training model in use which demands this type of commitment from the participants of a workshop. Without requests for release time, or for additional resources; participants were supposed to do evaluations of their own work, on their own time -- sometimes without support from their own institutions. The ATM workshop does not end after two weeks of good living and much talking; one is asked to make important professional commitments over a long period of time.

The organizers learned how difficult commitment is -- a long-term commitment, without incentives other than those of wanting to do a good job of whatever one is doing in one's working life. The organizers also learned that innovation (such as the ATM) can make the innovator vulnerable. The administrator playing the numbers game can ask: Why aren't there more of them? What did you do wrong that so many participants stayed away? Few would look at the other side of the coin and say that 33% of the trainees did return and that there were six evaluation studies at various stages of development.

Introduction of mixed cycles

It became quite clear during the January 1980 panel that it was unrealistic to expect a 100% rate of return from participants. Thirty per cent should be acceptable; 40% should be good; anything more than 50% very good indeed; and anything above 60% should be considered a miracle. To cope with low rates of return, the idea of mixed cycles was invented. We would begin with 20 people in the first workshop of cycle A, for instance; meet in a mid-term panel only those who have shown satisfactory progress on their evaluation studies; and in a second workshop of Cycle A, add new participants to bring the number of participants back to 20. The current structure of the ATM looks as follows:

A_1...(½ year)...P_a...(½ year)...A_2

↓

B_1...(½ year)...P_b...(½ year)...B_2

↓

C_1

...(½ year)...P_c...(½ year)...C_2

↓

D_1

Technical support is provided to participants throughout the year by the EFRC, as we have already indicated.

A_2/B_1: August 1980 workshop

The A_2/B_1 workshop took place in Kericho, Kenya, in August 1980 and was attended by 20 people -- 6 participants from Cycle "A" and 14 new participants joining the new Cycle "B".

The mid-term panel for the second cycle P_b took place during March 23-28, 1981 and was attended by 10 participants, 50% of the A_2/B_1 (August 80) participants. This was much better than the 32% rate of return of participants to the first cycle panel P_a.

B_2/C_1: June 1981 workshop

The third year cycle began with the B_2/C_1 workshop held during June 1-13, 1981. This workshop was attended by 3 participants from the A Cycle, 4 participants from the B Cycle and 7 new participants -- fourteen participants in all. (This might seem low as compared to 19 participants of A_1 workshop and 20 participants of A_2/B_1 workshop, but attendance had been kept deliberately low to provide better services to participants attending the workshop.

The mid-term panel of C Cycle (P_c) was conducted in February 22-27, 1982 and the C_2/C_1 workshop is now scheduled for June 28-July 10, 1982.

Productivity of Workshops

The rate for completion of evaluation studies as demonstrated by written reports of studies was 37% for Cycle A and 30.7% for Cycle B. However, it must be said that only one study out of Cycle A came close to the quality of studies produced in Cycle B. We will be able to judge the productivity of Cycle C later in July 1982.

An Important Criterion of Success:
Kenyans Take Over

The June-July 1982 (C_2/D_1) workshop is the last workshop which will receive outside technical assistance (in the form of Prof. H. S. Bhola as director of the workshop and Dr. Josef Muller as a member of the workshop faculty). The rest of the D Cycle and the few cycles to follow will be conducted by the Kenyan faculty independently, with only financial support from the German Foundation for International Development. This must be considered a sign of success.

In the meantime, DSE hopes to take the Action Training Model elsewhere in East, Central and Southern Africa to help where help is requested.

BIBLIOGRAPHY

We want to begin with a few comments on the usefulness of bibliographies in workshop documents of this type. Typically, bibliographies are included in books, monographs and research papers to demonstrate to readers that the piece of writing offered to them is part of a tradition of research and study. The bibliography shows the choices from the research tradition made by the author, indicating which works the author has considered important and useful and what new departures have been made or suggested. At the same time, a bibliography is an invitation to the reader to go independently to some of the books in the bibliography to further cultivate his or her own special interests.

It is the "invitation to the reader" aspect of this bibliography that deserves comment here. We realize, of course, that practitioners are, perhaps, not likely to be as enthusiastic readers as professors and researchers are supposed to be. (There are, of course, avid readers among practitioners and miserable nonreaders among professors and researchers!) More to the point, it is unlikely that most of the books in this bibliography will be part of the holdings of libraries to which most practitioners in the Third World will have access.

We have yet included this specially developed bibliography in this monograph not to show off but instead to indicate to the readers that training design, evaluation models and techniques, and evaluation of development training are important topics which are receiving increasing attention from researchers, planners and practitioners alike. It is our hope that this select bibliography will also serve as a catalog of issues and problems of training design and training evaluation that are today part of the concern of individual evaluators and developmental institutions. Lastly, it is our hope that our readers will want to get hold of some of the books and articles included here and will use their individual resources or the help of their librarian to obtain these materials to be read.

This bibliography is divided into three sections. The first section deals with the topic of *training design*. Some very selective books, and articles dealing with the various aspects of instructional and training design and training methodologies have been included in this section to give participants a sense of the "science of training design" that is becoming a reality. A second section includes books on *evaluation models and techniques*, covering both the scientific and the naturalistic paradigms. A third section deals with the *evaluation of training*, especially the evaluation of development training. Readers should note that evaluation needs have come to be recognized in almost all sectors of development -- agriculture, cooperatives, health, family planning, rehabilitation, media use and others.

Section I

Bibliography on Training Design

Boydell, T. H., *A Guide to the Identification of Training Needs*. London, U.K.: British Association for Commercial and Industrial Education, 1976.

Burrack, Elmer H., *Strategies for Manpower Planning and Programing*. Morristown, New Jersey: General Learning Press, 1972.

Caldwell, Lynton K., *Improving the Public Service Through Training*. Washington, D.C.: Agency for International Development, 1962.

Cookey, Samual J., "The Training and Supply of Middle-Level Personnel," *Prospectus*, VI, 2:223-230.

Craig, Robert L. and Lester R. Bittel (Eds.), *Training and Development Handbook*. New York, N.Y.: McGraw-Hill, 1967.

Crawford, Meredith P., "Concept of Training," in Robert M. Gagne (Ed.), Psychological Principles in System Development. New York, N.Y.: Holt, Rinehart and Winston, 1963.

Davies, Ivor K., Competency Based Learning: Technology, Management, and Design. New York, N.Y.: McGraw-Hill, 1973.

deClerck, Marcel, The Operational Seminar: A Pioneering Method of Training for Development. Paris: Unesco, 1976.

Denisi, Angelo S., "The Implications of Job Clustering for Training Programs," in Journal of Occupational Psychology, 49, 1976.

Dick, W. and L. Carey, The Systematic Design of Instruction. Glenview, Ill.: Scott, Foresman and Co., 1978.

Ely, Donald D., "Training by Objectives, A Systems Approach to Instruction," Training and Development Journal, 29, June 1975.

Gilbert, Thomas F., "Praxeonomy: A Systematic Approach to Identifying Training Needs," Management of Personnel Quarterly, Fall 1967.

Gretlar, Armin, The Training of Adult Middle Level Personnel. Paris: Unesco, 1972.

Harbison, Frederick, Human Resources as a Wealth of Nations. New York, N.Y.: Oxford University Press, 1973.

Harris, Ben M., Improving Staff Performance Through In-service Education. Allyn and Bacon, 1973.

Havelock, R. G. and Marcy C. Havelock, Training for Change Agents: A Guide to the Design of Training Programs in Education and Other Fields. Ann Arbor, Michigan: Center for Research on Utilization of Scientific Knowledge, University of Michigan, 1973.

Kent, George, "Training is a System," Training and Development Journal, Vol. 29, September 1975, pp. 10-12.

Laird, D., Approaches to Training and Development. Reading, Mass.: Addison-Wesley Pub. Co., 1978.

Mager, R. F., Analyzing Performance Problems. Belmont, Ca.: Fearon Publishers, 1970.

Nadler, L., "Organizational Factors which Affect the Accountability of Training Effort," Training and Development Journal, Vol. 25, No. 10, pp. 6-8.

Odiorne, George S., Training by Objectives. New York, N.Y.: The Macmillan Co., 1970.

Otto, Calvin P. and Rollin O. Glaser, The Management of Training. Reading, Mass.: Addison Wesley, 1970.

Schaffer, B. B. (Ed.), Administrative Training and Development. New York, N.Y.: Praeger, 1974.

Stockard, James G., Career Development and Job Training. New York, N.Y.: Amacom, 1977.

Tracey, William R., Designing Training and Development Systems. American Management Association, 1971.

Welch, J., A Task Unit Concept for On-the-Job Training. Columbia, Miss.: University of Missouri, 1966.

Section II

Bibliography on Evaluation

Models and Techniques

Anderson, D. and P. Banjaminson, Investigative Reporting. Bloomington, Ind.: I.U. Press, 1976.

Anderson, Scarvia B., Samuel Ball, Richard T. Murphy and Associates, Encyclopedia of Educational Evaluation: Concepts and Techniques for Evaluating Education and Training Programs. San Francisco, Ca.: Jossey-Bass, 1975.

Babbie, Earl R., Survey Research Methods. Belmont, Ca.: Wadsworth Publishing Co., 1973.

Banaka, W. H., Training in Depth Interviewing. New York, N.Y.: Harper and Row, 1971.

Barker, R. G. and Associates, Habitats, Environments, and Human Behavior: Studies in Ecological Psychology and Eco-Behavioral Science. San Francisco, Jossey-Bass, 1978.

Becker, H. S., Sociological Work: Methods and Substance. Chicago: Aldine, 1970.

Bickman, L., and T. Henchy (Eds.), Beyond the Laboratory: Field Research in Social Psychology, New York, N.Y.: McGraw-Hill, 1972.

Bloom, Benjamin S., and others, Taxonomy of Educational Objectives - Handbook I: Cognitive Domain. New York, N.Y.: David McKay, 1975.

Bogdan, Robert, Participant Observation in Organizational Settings. New York, N.Y.: Syracuse University Press, 1972.

Bogdan, Robert, and Steven J. Taylor, Introduction to Qualitative Research Methods. New York, N.Y.: John Wiley and Sons, 1975.

Brady, J., *The Craft of Interviewing*. Cincinnati, Ohio: Writer's Digest, 1976.

Campbell, Donald T. and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. Chicago, Ill.: Rand McNally, 1966.

Conner, Ross F. (Ed.), *Methodological Advances in Evaluation Research*. Beverley Hills, Ca.: Sage, 1981.

Cook, Thomas D. and Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, Ill.: Rand McNally, 1966.

Cook, Thomas D. and Charles S. Reichardt, *Qualitative and Quantitative Methods in Evaluation Research*. Beverley Hills, Ca.: Sage, 1979.

Dexter, L. A., *Elite and Specialized Interviewing*. Evanston, Ill.: Northwestern University Press, 1970.

Doughlas, Jack D., *Investigative Social Research: Individual and Team Field Research*. Beverley Hills, Ca.: Sage, 1976.

Eisner, E. W., *The Educational Imagination*. New York, N.Y.: Macmillan, 1979.

Freeman, Howard E. and Clarence C. Sherwood, *Social Research and Social Policy*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Glaser, Barney G, and Anselm L. Strauss, *The Discovery of Grounded Theory*, Chicago, Ill.: Aldine Publishing Company, 1967.

Glass, Gene V. and Julian C. Stanley, *Statistical Methods in Education and Psychology*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.

Gronlund, Norman E., *Preparing Criterion-Referenced Tests for Classroom Instruction*. New York, N.Y.: Macmillan Co., 1973.

Guba, Egon G. and Yvonna S. Lincoln, *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*. San Francisco, Ca.: Jossey-Bass, Inc., 1981.

Henerson, Morris and Fitz-Gibbon. *How to Measure Attitudes*. Beverly Hills, Ca.: Sage, 1978.

Holsti, Ole R., *Content Analysis for the Social Sciences and Humanities*. Reading, Mass.: Addison-Wesley Publishing Co., 1969.

Jacobs, G. (Ed.), *The Participant Observer*. New York, N.Y.: Braziller, 1970.

Johnson, John M., *Doing Field Research*. New York, N.Y.: The Free Press, 1975.

Joint Committee on Standards for Educational Evaluation, Standards for Evaluations of Educational Programs, Projects, and Materials. New York, N.Y.: McGraw-Hill, 1981.

Kaplan, A., The Conduct of Inquiry. San Francisco, Ca.: Chandler, 1964.

Kerlinger, Fred N., Foundations of Behavioral Research (2nd Ed.). New York, N.Y.: Holt, Rinehart and Winston, Inc., 1973.

Krathwohl, David R., and others, Taxonomy of Educational Objectives - Handbook II: Affective Domain. New York, N.Y.: David McKay, 1974.

Krippendorff, Klaus, Content Analysis: An Introduction to Its Methodology. Beverley Hills, Ca.: Sage, 1980.

Lofland, J., Analyzing Social Settings. Belmont, Ca.: Wadsworth, 1973.

Marshall, Jon Clark and Loyde Wesley Hales, Essentials of Testing. Reading, Mass.: Addison-Wesley Publishing Co., 1972.

Morris, Lynn Lyons, Carol Taylor, Fitz-Gibbon and Marlene E. Henerson, Program Evaluation Kit: (1) Evaluator's Handbook; (2) How to Deal with Goals and Objectives; (3) How to Design Program Evaluation; (4) How to Measure Program Implementation; (5) How to Measure Attitudes; (6) How to Measure Achievement; (7) How to Calculate Statistics; and (8) How to Present an Evaluation Report. Beverley Hills, Ca.: Sage, 1978.

Oppenheim, A. N., Questionnaire Design and Attitude Measurement. New York, N.Y.: Basic Books, 1966.

Parlett, M. and G. Dearden (Eds.), Introduction to Illuminative Evaluation: Studies in Higher Education. Beverley Hills, Ca.: Sage, 1978.

Patton, M. Q., Utilization-Focussed Evaluation. Beverley Hills, Ca.: Sage, 1978.

Payne, S. J., The Art of Asking Questions. Princeton, N.J.: Princeton University Press, 1951.

PDK National Study Committee on Evaluation, Educational Evaluation and Decision Making. Itasca, Ill.: F. E. Peacock Publishers, 1971.

Popham, W. James (Ed.), Evaluation in Education: Current Applications. Berkeley, Ca.: McCutchan, 1974.

Provus, Malcolm, Discrepancy Evaluation for Educational Program Improvement and Assessment. Berkeley, Ca.: McCutchan, 1971.

Rich, J., Interviewing Children and Adolescents. New York, N.Y.: St. Martin's Press, 1968.

Richardson, S. A., B. S. Dohrenwend and D. Klein, Interviewing: Its Forms and Functions. New York, N.Y.: Basic Books, 1965.

Rippey, R. M. (Ed.), Studies in Transactional Evaluation. Berkeley, Ca.: McCutchan, 1973.

Sanders, W. B., The Sociologist as Detective: An Introduction to Research Methods. (2nd ed.). New York, N.Y.: Praeger, 1976.

Saxe, Leonard, and Michelle Fine, Social Experiments: Methods for Design and Evaluation. Beverley Hills, Ca.: Sage, 1981.

Scriven, M., "Maximizing the Power of Causal Investigations: The Modus Operandi Method." IN W. J. Popham (Ed.), Evaluation in Education. Berkeley, Ca.: McCutchan, 1974.

Scriven, M., "The Methodology of Evaluation." AERA Monograph Series in Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.

Siegel, S., Nonparametric Statistics for Behavioral Sciences. New York, N.Y.: McGraw-Hill, 1956.

Slonim, Morris James, Sampling. New York, N.Y.: Simon Schuster, 1960.

Spradley, J. P., The Ethnographic Interview. New York, N.Y.: Holt, Rinehart and Winston, 1979.

Spradley, J. P., Participant Observation. New York, N.Y.: Holt, Rinehart and Winston, 1980.

Stake, R. E. and J. A. Easley, Jr. (Eds.), Case Studies in Science Education. Urbana, Ill.: Center for Instructional Research and Curriculum Evaluation, University of Illinois, 1978.

Thompson, Mark, Benefit-Cost for Program Evaluation. Beverley Hills, Ca.: Sage, 1980.

Turner, R. (Ed.), Ethnomethodology. Middlesex, England: Penguin Books, 1974.

Tyler, R. W., Constructing Achievement Tests. Columbus, Ohio: Ohio State University Press, 1934.

Tymitz, B. and R. L. Wolf, An Introduction to Judicial Evaluation and Natural Inquiry. Washington, D.C.: Nero and Associates, 1977.

Wax, R., Doing Fieldwork: Warnings and Advice. Chicago: University of Chicago Press, 1971.

Webb, E. J., D. T. Campbell, R. D. Schwartz and L. Sechrest, Unobtrusive Measures. Chicago: Rand McNally, 1966.

Weisberg, H. F. and B. D. Bowen, An Introduction to Survey Research and Analysis. San Francisco, Ca.: W. H. Freeman, 1977.

Weiss, Carol H., Evaluating Action Programs: Readings in Social Action. Boston, Mass.: Allyn and Bacon, 1972.

Williams, P. N., Investigative Reporting and Editing. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Willis, G. (Ed.), Qualitative Evaluation. Berkeley, Ca.: McCutchan, 1978.

Worthen, Blaine R. and James R. Sanders. Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones, 1973.

Section III

Bibliography on Evaluation of Training and Development

Apthorpe, Raymond, "Some Problems of Evaluation," P. 209-229, in Cooperatives and Rural Development in East Africa (Ed. Carl Gosta Widstrand). Uppsala: The Scandinavian Institute for African Studies, 1970, 217 pages.

Armenakis, Achilles A. and others, "Evaluation Guidelines for the OD Practitioner," Personnel Journal, Vol. 54, No. 2, February 1975, p. 102.

Bowers, D. G. and J. L. Franklin, Survey Guided Development: Data Based Organizational Change, University of Michigan, University of Michigan Center for Research on Utilization of Scientific Knowledge (CRUSK), Institute for Social Research, Organizational Development Research Program, 1976, 166 pages.

Byrn, Darcie, (Ed.), Evaluation in Extension, prepared by the Division of Extension Research and Training, Federal Extension Service, U.S. Department of Agriculture, Topeka, Kansas, 1967, 107 pages.

Cantalanello, R. F. and D. L. Kirkpatrick, "Evaluating Training Programs -- The State of the Art," Training and Development Journal, May 1968, pp. 2-9.

Chabotar, Kent T. and Lawrence T. Lad, Evaluation Guidelines for Training Programs. Midwest Intergovernmental Training Committee, Ann Arbor, Michigan State Department of Civil Service, 1974.

Cronback, Lee J., "Evaluation for Course Improvement," Teachers College Record, 64, 1963.

Dubois, Philip H., and E. Douglas Mayer (Eds.), Research Strategies for Evaluating Training, AERA Monograph Series, No. 4, Chicago: Rand McNally, 1970.

Frederiksen, Norman et al., Prediction of Organizational Behavior. New York, N.Y.: Pergamon Press, Inc., 1972, 333 pages.

Freedman, Ronald and John Y. Takeshita, Family Planning in Taiwan: An Experiment in Social Change. Princeton, N.J.: Princeton University Press, 1969, 500 pages.

Ghorpade, Jaisingh, Assessment of Organizational Effectiveness: Issues, Analysis, Readings. Pacific Palisades, Ca.: Goodyear Publishing Co., Inc., 1971, 256 pages.

Grotelueschen, Arden D., and others, Evaluation in Adult Basic Education: How and Why. Danville, Ill.: Interstate Printers, 1976, 274 pages.

Hamblin, A. C., Evaluation and Control of Training. London, U.K.: McGraw-Hill, 1974.

Hefferin, E. and A. Katz, "Issues and Orientations in the Evaluation of Rehabilitation Programs: A Review Article. Part One." Rehabilitation Literature, 1971, 32, 66-75. Part Two, 98-107.

Jacque, J., Manual on Cooperative Management. Manchester: Cooperative Union, 1976, 184 pages.

Katz, F. M., Guidelines for Evaluating a Training Programme for Health Personnel. Switzerland: WHO, 1978.

Kinyanjui, Peter E., "Radio Correspondence Courses in Kenya: An Evaluation," Educational Broadcasting International, Vol. 6, No. 4 (1973), 180-188.

Lathan, Michaiel C., Planning and Evaluation of Applied Nutrition Programmes. Rome: FAO (FAO Nutritional Studies No. 26), 1972, 125 pages.

Lecht, Leonard A., Evaluating Vocational Education: Policies and Plans for the 1970's with an Annotated Bibliography. New York: Praeger Publishers, 194 pages.

Marguilies, Newton, and Anthony Raia, Organizational Development: Values, Process and Technology. New York, N.Y.: McGraw-Hill Book Co., 1972, 640 pages.

Mauldin, W. P., "Measurement and Evaluation of National Family Planning Programs," Demography, 1967, 4, 71-80.

Mbakile, E. P. R., Evaluation Report on the First Phase of the Radio Education Programme. Mwanza, Literacy Project, 1974, 24 pages.

McLaughlin, Stephen and Jeanne Moulton, Evaluating Performance Training Methods: A Manual for Teacher Trainers, 1975.

Mitchell, E. F., Cooperative Vocational Education: Principles, Methods and Problems. Boston: Allyn & Bacon, Inc., 1977, 398 pages.

Picket, Liam E., The Use of Statistics: A Handbook for Cooperatives, ILO, 1975.

Price, James L., Handbook of Organizational Measurement. Lexington, Mass.: D. C. Heath & Co., 1972, 210 pages.

Roemer, Milton I., Evaluation of Community Health Centers. Geneva: World Health Organization (WHO), 1972, 42 pages.

Rogers, Everett and Rekha Agarwala-Rogers, Evaluation Research on Family Planning Communication. Boudin, Paris: Unesco, 1976, Population Division, Population Communications: Technical Document No. 4, 127 pages.

Schmuck, Richard and A. Runkel, Handbook of Organizational Development in Schools. Palo Alto, Ca.: National Press Books, 1972.

Weiss, Carol H., Evaluation Action Programs: Reading in Social Action and Education. Boston: Allyn and Bacon, Inc., 1972, 365 pages.

WHO Regional Office for Europe, Methods of Evaluating Public Health Programmes. (Report of a Symposium, Keil, 14-17 November 1967), Copenhagen, 1968.

GLOSSARY OF TERMS

Action training model (ATM). A training model developed under the aegis of the German Foundation for International Development (DSE), Bonn, Federal Republic of Germany. The model emerged within the context of a series of workshops on the evaluation of basic education and development training programs. The model is so called for its emphasis on action. Trainees are required to make commitments to a full cycle of training experiences: first, a workshop where trainees learn generally about evaluation and develop specific proposals for evaluation studies; second, a mid-term panel where the trainees come with evaluation data collected by them during some six months of the implementation of their evaluation studies, review their experiences and prepare for data analysis; and, finally, another workshop where old trainees come back to report on their findings and new ones launch upon a new training cycle under the ATM.

Analysis of variance (ANOVA). A method of determining whether the differences between groups are statistically significant.

Attrition. Loss of subjects from a chosen sample during the course of a study.

Audit of an evaluation. Examination and verification by another independent team of the quality of an evaluation plan, the adequacy with which it was implemented, the accuracy of its findings, and the validity of its conclusions.

Base-line survey. An initial survey that can serve as a base for comparing changes observed subsequently.

Bias. A consistent alignment to one particular point of view which may make objective evaluation results improbable.

Case study. A detailed description and analysis of a single program, project, course or an instructional material conducted within its educational or social context.

Code. To convert a given set of data or items into a set of quantitative or qualitative symbols. (Examples: 1, 2, 3 and 4; L, M and H.)

Coefficient. A statistic (or value) that represents the degree of occurrence of a property or relationship. (Example: correlation coefficient.)

Concept analysis. The process of "unpacking" concepts to define them with such precision that they will have the same invariant meanings for all readers.

Content analysis. Identifying, categorizing and listing according to some rules, ideas, references, feelings or judgments found in a set of transcripts, documents, etc.

Context evaluation. Assessing and evaluating the environmental variables of a program.

Control group. A group which resembles an experimental group (the group which is subjected to a particular program or method) as closely as possible, but is not exposed to the program or method whose effect is being studied. It thus serves the comparative purposes of the evaluator.

Correlation. A statistic which indicates the degree of relationship (going together or happening together) between or among variables. Correlations can vary from -1.00 to +1.00.

Cost-benefit analysis. An assessment of the inputs and outputs of a program in terms of their monetary values.

Cost-effectiveness analysis. An assessment of the inputs, processes and outputs of a program in terms of the effectiveness of means employed for the ends obtained.

Criterion. A standard by which something is judged.

Criterion-referenced tests. Tests whose scores are interpreted according to the criteria of performance specifically defined by the teacher in regard to a particular group, and not by reference to performance of some comparable populations.

Data. Material gathered during the course of an evaluation study (both quantitative and qualitative) which is then used to develop information for decision making.

Data analysis. The process of identifying ideas, themes, and hypotheses from the data, and the use of data to demonstrate support for them.

Data pieces. Individual tests, interview schedules, questionnaires and diaries that have been completed as part of the data collection phase of an evaluation study.

Dependent variable. A measure (for example, better nutritional habits) which is supposed to vary as a result of the introduction of an independent variable (for example, teaching of nutritional habits by the family health educator).

Design. A model or a clearly established set of procedures to determine how an evaluation study will be conducted. (Also see training design.)

Development. The processes that lead to greater production of wealth in a society and a just and equitable distribution of such wealth, accompanied by progressive consumption of education and culture, and commitments to universal brotherhood, peace and preservation of the globe.

Development training. Training of workers and change agents who will, in turn, impart economic, social and political skills to farmers, workers, housewives and youth to enable them to generate and sustain development within their societies.

Dissemination. The process of spreading information about evaluation objectives and results among those concerned with the evaluation study. The methods of dissemination may be written or oral.

Evaluation. Objective and systematic collection of information about a program, project, or instructional material for its improvement. (More recently in literature, evaluation is being defined as the "systematic investigation of the worth or merit of an object; e.g., a program, project, or instructional material.)

Evaluation system. An arrangement of methods, procedures and plans of action designed to provide decision-makers with information on the inputs, outputs, context and process of a given program.

External evaluation. Evaluation conducted by evaluators not on the staff of a program or project.

Extrapolate. To infer from what is known, something that is unknown. (Population figures for a country for the year 2000, may be extrapolated from the population growth figures during 1950-80.)

Feedback. A term borrowed from electronics: the return of part of the output of a system into the input for purposes of modification and control of the output. In the context of program planning, feedback means evaluative information on program effects.

Field test. A preliminary study of a program, project or instructional material in a setting very similar to the one in which it will be later implemented or used on a much larger scale.

Formative evaluation. Evaluation conducted during the very formation of a program, project or instructional material.

Generalizability. The extent to which claims and assertions made about a program, project or instructional material in one setting can be applied in other settings.

Goal-free evaluation. Evaluation of outcomes of programs and projects where the evaluator functions without knowledge of the purposes and goals of a program or project.

Human resource development (HRD). The education and training of manpower, both for formal and informal sectors of the economy, using both formal and nonformal systems of instruction.

Independent variable. A treatment variable introduced in an evaluation setting (example: a new teaching method), anticipated to create varying effects on a dependent variable (for example, performance on a test).

Indicator. Something that indicates, points, signifies; a gauge that represents another entity. Thus, a high drop-out rate in an adult education program may be an indicator of a lack of community motivation.

Input evaluation. Assessing the various resources used in conducting a program.

Institution building. The process of developing organizational arrangements or systems for the implementation of programs or projects on a long-term basis. (To institutionalize is to make a program more or less permanent through institution building.)

Instrument. An assessment device (test, questionnaire, interview schedule, or observation schedule) used for the purposes of evaluation.

Internal evaluation. An evaluation conducted by a staff member from within the organization whose program, project or instructional material is being evaluated.

Level of significance. A predetermined probability value which is used to decide whether the results of an evaluation study occurred by chance and were not really a consequence of a program, project or instructional material. ($p = < .01$ means that there is the probability of only one in a hundred for the program effect to have appeared by chance.)

Management information system (MIS). A system (computorized, manual or a mix of the computorized and manual) including planning and implementation data in regard to a program or project. (See also Monitoring.)

Matching. The process by which subjects assigned to different groups are made to be as equivalent as possible. (Matching may be done on such variables as sex, age, education, socioeconomic status, etc. A set of twins would be perfectly matched for purposes of some studies.)

Mean. The sum of a group of scores divided by the number of scores.

Median. The score in a group of scores that is midway in the distribution.

Mode. The score in a group of scores that occurs most often.

Model. A design, description or analogy used to help visualize or make understandable something that is more complex.

Modus operandi analysis. A procedure similar to detective work whereby causes and effects are hypothesized, tested and analyzed to arrive at the most likely patterns of events and their consequences.

Monitoring. To monitor is to check on an on-going program or project for flaws or breakdowns, to enable decision-makers to regulate activities and to undertake corrective action. Monitoring is typically based in a management information system.

Naturalistic inquiry paradigm. Study of behavioral phenomena in natural settings and in their normal context, using methods drawn from ethnography, anthropology and sociological field studies. Also called the ethnographic or the phenomenological paradigm.

Needs assessment. The process of ascertaining the learning needs, health needs or other developmental needs of beneficiaries of educational and developmental programs. Needs assessments are a mix of "felt" needs expressed by beneficiaries and new needs "fashioned" by change agents

No significant difference (nsd). A label which is used to say that the observed difference between two statistics could have occurred by chance. (See Level of confidence above.)

Nonformal education. A collection of organized or semi-organized educational activities, operating outside the formal education system and meeting the immediate educational needs of both conventional and non-conventional learners. (Formal education is that which is provided by schools, colleges and universities. Informal education is that where neither the educator nor the one being educated are conscious of the process of teaching-learning taking place.)

Norm. A value or pattern of values representing the typical performance of a group or population.

Norm-referenced tests. See standardized tests.

Objectives-referenced tests. Tests whose scores are interpreted according to the objectives which a program project or course was designed to teach, without comparing performance of other groups on the test.

Operational seminar. A training method developed within Unesco wherein participants experience on a reduced time-scale the total process of community work, problem diagnosis, needs assessment, field organization, materials design and evaluation in an actual field setting.

Output evaluation. Assessing the quality and quantity of the final product(s) of the program, also taking into account any unintended by-products of the program.

Paradigm. An example or pattern; a very clear example of an archetype. In evaluation, a paradigm is equivalent to the 'intellectual ideology' of an evaluator.

Parameter. Any one of a set of properties whose value determines the characteristics or behavior of something.

Participative approaches. Designs, procedures and methods of planning, implementation and evaluation that are built upon the active involvement of the would-be beneficiaries of programs and projects.

Population. All the persons in the group to which the results from a study will apply. (E.g., all cotton farmers in the Lake Regions of Tanzania, all women in the child-bearing age in Indiana.)

Post-test. A test to determine the effects of a program, project or instructional material after application or completion.

Pre-test. A test to determine level of performance before the start or application of a program, project or instructional material.

Problem complex. A whole set of interrelated problems (of planning, or of management, or of evaluation), emerging around a decision point within a system.

Process evaluation. Assessing procedural strategies and comparing effectiveness of different approaches to instruction, extension, animation and organization.

Product evaluation. Assessing the effectiveness of curricular or instructional products.

Qualitative data. Facts, claims and assertions in narrative form, and not in numbers. (Qualitative data can, however, be converted into numerical form by coding and scoring.)

Quantitative data. Facts, claims and assertions presented in numerical forms.

Quick appraisals. Quick evaluations, less comprehensive and less exhaustive than regular evaluations, conducted under conditions of emergency to investigate the cause of a breakdown, to anticipate problems or to get early returns on the impact of a program.

Random sample. A representative portion chosen from among the population; each individual in the population has an equal chance of being selected each time a selection is made.

Reliability. The property of an instrument giving the same reading or score when used by different investigators on the same entity, or by the same investigator repeatedly on the same one entity.

Replication. The repeat of an evaluation study with all essential aspects of the study remaining unchanged.

Responsive evaluation. Evaluation that responds to the information needs of the various stakeholders in a program by providing evaluation feedback on concerns and issues raised by them, rather than evaluating what the evaluator thinks is worth evaluating.

Sample. A part of a population chosen according to some method to represent the total population.

Scientific inquiry paradigm. The approach borrowed from the hard sciences involving experimental design, randomized samples, controlled groups and statistical analysis.

Situation-specific strategy (3-S) model of evaluation. A five-step model that relates evaluation with change; requires the articulation of means and ends within an educational or a developmental program; proposes the development of profiles of information needs; suggests that situation-specific and strategic agendas for evaluation be developed; and that the choice of evaluation methodologies and techniques be both technically appropriate and practically feasible within the setting of evaluation.

Standard deviation (s). A measure of variability calculated on the basis of differences of individual scores within a group from the group mean. (s^2 is called variance.)

Standardized tests. Tests whose scores are interpreted in comparison with some norms established in terms of some larger groups or populations.

Statistic. A summary number that describes the characteristic or property of a sample.

Statistical analysis. An examination of complex relationships between variables using empirical data and rules of statistics.

Statistics. The science of methods for analyzing data obtained from empirical observations to make descriptions or inferences. Thus, there is descriptive statistics, and there is inferential statistics.

Summative evaluation. Assessment of the impact of the total product, program, etc., comparing observed effects with anticipated or desired effects.

System. A whole emerging from an interacting and interdependent set of parts, subject to a common plan and having a common purpose.

Systems model. A model that looks at social reality as a system that can always be described in terms of inputs, processes, outputs and context. (See also model and system.)

Taxonomy. An orderly classification that has some theoretical underpinnings.

Thick description. Detailed and faithful descriptions in the form of photographic records and protocols or written case studies.

Training design. A model or a clearly established set of procedures to develop a training program, involving planful selection of educational objectives, learner characteristics, teaching methodologies and learning environments.

Triangulation. Comparing and testing results from two or more different approaches to the solution of the same one problem.

Unit of analysis. The social unit such as individual, husband-wife dyad, family, group, organization or community which is the focus of interest for the evaluator; which will determine the organization of data; and about whose behavior statements, claims and assertions will be made.

Unobtrusive measures. Methods of examination in which the evaluators do not materially interfere in the situation, but rely on indirect procedures to gather data.

Validity. The property of an instrument to be able to measure what it was supposed to measure.

Variable. A characteristic that can take on different values.

Variance. A measure of variability calculated on the basis of differences of individual scores within a group from the group mean. The square root of variance gives the value of standard deviation (s).